Monardo, Julien

Thesis

Essays on structural demand estimation : = Essais sur lestimation structurelle de la demande

Essais sur l'estimation structurelle de la demande

Julien Monardo

NNT : 2019SACLN042

Thèse de doctorat

Essays on Structural Demand Estimation

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'École normale supérieure Paris-Saclay

Ecole doctorale n°578 Sciences de l'homme et de la société (SHS)
Spécialité de doctorat : Sciences économiques

Thèse présentée et soutenue à Cachan, le 18 octobre 2019, par

Julien Monardo

Composition du Jury :

Debopam Bhattacharya Professeur associé, University of Cambridge	Rapporteur
Alfred Galichon Professeur, New York University	Rapporteur
Steven Berry Professeur, Yale University	Président
Xavier d'Hautfœuille Administrateur Hors Classe, INSEE (CREST)	Examinateur
Laura Grigolon Professeur assistant, University of Mannheim	Examinateur
André de Palma Professeur des universités, ENS Paris-Saclay (CREST)	Directeur de thèse

Remerciements

J'aimerais commencer cette thèse en consacrant quelques mots aux personnes sans lesquelles cette thèse n'aurait pas été possible.

Mes premiers remerciements vont à mon directeur de thèse, André de Palma, pour ses conseils et son soutien durant ces trois années. Je lui suis très reconnaissant de m'avoir fait bénéficier de son expertise et de son expérience, notamment à travers la co-écriture du premier chapitre de cette thèse. Les nombreuses discussions que nous avons eues m'ont beaucoup appris.

Je tiens également à remercier tous les membres de mon jury : Debopam Bhattacharya et Alfred Galichon d'avoir accepté la charge de rapporteur ; ainsi que Steven Berry, Xavier d'Hautfœuille et Laura Grigolon d'avoir accepté d'être membres de mon jury.

Je souhaiterais également remercier toutes les personnes avec lesquelles j'ai eu la chance de travailler. Tout d'abord, Alessandro Iaria qui m'a initié à l'économie industrielle empirique à l'ENSAE et au CREST et qui m'a donné envie de faire cette thèse ; ensuite, Mogens Fosgerau avec qui j'ai pris plaisir à co-écrire et pour son accueil chaleureux durant mes deux séjours à l'université de Copenhague le premier chapitre de cette thèse ; Ao Wang avec qui j'ai partagé de nombreuses discussions passionnantes ; et Nathalie Picard avec laquelle je partage un projet en économie de la famille.

Je suis tout particulièrement reconnaissant envers Morgane Cure pour son soutien, ses conseils et sa bienveillance, et envers Hugo Molina pour son soutien, ses conseils et les conversations interminables que nous avons eues. Leur présence au cours de ces trois dernières années a été précieuse.

Je souhaiterais ensuite remercier ceux avec qui j'ai eu l'honneur de partager mes bureaux à l'ENS Paris-Saclay et au CREST : Bastien Alvarez, Rémi Avignon, Florian Bonnet, Etienne Chamayou, Morgane Guignard, Juan Daniel Hernandez, Alessandro Ispano, Manuel Marfan Sanchez, Emmanuel Paroissien, Marine Sales, Benjamin Walter, Jiekai Zhang.

Je remercie également toutes les personnes de l'ENS Paris-Saclay et du CREST avec lesquelles j'ai partagé ces trois dernières années, les chercheurs, doctorants, stagiaires, et membres du personnel administratif. En particulier, je remercie Laurent Linnemer et Emmanuelle Taugourdeau pour la bienveillance dont ils ont fait preuve à mon égard. Je souhaite en particulier remercier Marie-Laure Allain, Christophe Bellego, Marie-Laure Cabon-Dhersin, Arthur Cazaubiel, Claire

Chambolle, Philippe Choné, Eva Drabik-Blervaque, Nicolas Drouhin, Nathalie Etchart-Vincent, Roxana Fernandez, Hubert Kempf, Francis Kramarz, Sébastien Laffitte, Alexis Larousse, Morgane Le Breton, Samuel Ligonnière, François Pannequin, Jean-Christophe Tavanti, Farid Toubal, Thomas Vendryes, Thibaud Vergé.

Ma thèse a également bénéficié des discussions avec plusieurs doctorants et chercheurs, en particulier Steven Berry, Xavier d'Hautfœuille, Yannick Guyonvarch, Olivier de Groote, Ali Horstascu, Jonas Lieber, Laurent Linnemer et Aureo de Paula.

Je remercie également mes collègues de Telecom ParisTech qui m'ont accueilli en post-doc au sein du département SES, notamment Marc Bourreau, Lukasz Grzybowski et Christine Zulehner. J'en profite aussi pour remercier Marie-Laure Allain, Roxana Fernandez, Xavier d'Hautfœuille et Farid Toubal pour leur aide précieuse dans l'obtention de ce post-doc.

Enfin, je témoigne toute ma reconnaissance à ma famille et mes amis (hors recherche), en particulier Aurélie Monardo, Florence Monardo, Claire Rougevin-Baville, Louis de Catheu, Guillaume Pommey et Ambroise Pouzoulet.

Résumé de la thèse

L'estimation structurelle des modèles de demande sur des marchés de produits différenciés joue un rôle important en économie. Elle permet de mieux comprendre les choix des consommateurs (e.g., en estimant les élasticités-prix de la demande). De plus, elle est le point de départ de l'étude de plusieurs questions économiques d'intérêt, incluant celles relatives au pouvoir de marché des entreprises (Berry *et al.*, 1995; Nevo, 2001), à la fusion d'entreprises (Nevo, 2000), à l'introduction de nouveaux produits sur le marché (Petrin, 2002; Gentzkow, 2007), à la politique commerciale (Goldberg, 1995; Verboven, 1996; Berry *et al.*, 1999), et aux taxes (Griffith *et al.*, 2019).

La littérature théorique a mis en évidence que les réponses à ces questions dépendent de la forme de la fonction de demande, laquelle est décrite par sa pente et sa courbature. Ainsi, étant donné un modèle d'offre (e.g., modèle statique de concurrence oligopolistique en prix), la qualité des réponses repose sur la capacité du modèle de demande à être "flexible", i.e., sur sa capacité à capter de manière flexible les substitutions qui existent entre les produits.

L'approche standard consiste à spécifier un modèle d'utilité aléatoire additif, à en calculer sa fonction de demande, et à estimer cette dernière en utilisant la méthode développée par Berry (1994). Le modèle d'utilité aléatoire additif est utilisé pour sa capacité à modéliser le comportement de consommateurs hétérogènes choisissant parmi un grand nombre de produits différenciés de manière parcimonieuse et flexible. La méthode de Berry (1994) est utilisée pour estimer des modèles de demande pour des produits qui sont différenciés de manières observée et inobservée par le modélisateur. Elle résout les problèmes d'endogénéité liés à la présence de termes d'erreurs structurels, lesquels représentent les caractérististiques des produits qui sont inobservées par le modélisateur mais observées et valorisées par les entreprises et les consommateurs. Elle consiste à estimer les paramètres structurels de la fonction de demande à partir du système d'équations qui égalise les demandes observées aux demandes prédites par le modèle. Or, les termes structurels d'erreurs entrent dans ce système de manière non-linéaire, empêchant donc l'utilisation des techniques standards des variables instrumentales. Berry (1994) propose ainsi d'inverser le système afin d'obtenir des équations de demande inverse au sein desquelles les termes d'erreurs structurels entrent de manière linéaire et de les utiliser comme base pour l'estimation. Toutefois, en général, ces demandes inverses n'ont pas

d'expression analytique. L'inversion doit donc être faite numériquement, ce qui exige l'emploi de procédures d'estimation non-linéaire et la résolution de problèmes connexes d'optima locaux et de précision de l'inversion numérique (Knittel & Metaxoglou, 2014).[1]

La méthode de Berry *et al.* (1995), connue sous le nom de méthode BLP, est la méthode la plus populaire et la plus avancée de cette approche. Elle utilise un modèle logit à coefficients aléatoires qu'elle estime par un algorithme réalisant une inversion numérique de la demande, imbriquée dans une procédure d'estimation non-lineaire. Elle permet de capter de manière flexible les substitutions entre les produits tout en résolvant les problèmes d'endogénéité. Toutefois, elle est sujette à des difficultés pratiques : la flexibilité exige l'utilisation de nombreux coefficients qui peuvent être difficiles à identifier empiriquement ; de plus, estimer un modèle logit à coefficients aléatoires peut être difficile et chronophage puisque cela exige l'emploi de procédures d'estimation non-linéaire ainsi que la simulation et l'inversion numérique des fonctions de demande.

L'autre méthode très répandue utilise le modèle logit emboîté, lequel évite les difficultés associées à la méthode BLP en ayant uniquement recours à des régressions linéaires. Toutefois, le modèle logit emboîté est critiqué au motif qu'il ne permet pas de capter de manière flexible les substitutions entre les produits et qu'il demande au modélisateur de définir la structure des nids avant l'estimation, i.e., de déterminer les sources pertinentes de segmentation du marché.

Cette thèse poursuit l'objectif de proposer des modèles de choix des consommateurs qui soient flexibles et qui aboutissent à des méthodes d'estimation simples et rapides. Pour cela, elle adopte une approche différente : elle développe de nouveaux modèles de demande inverse qui sont cohérents avec un modèle d'utilité de consommateurs hétérogènes. Cette approche permet de capter de façon flexible les substitutions entre les produits, grâce à de simples régressions linéaires basées sur des données incluant les parts de marché, les prix et les caractéristiques des produits. Ces modèles peuvent être utilisés dans différents domaines de l'économie, incluant l'économie industrielle, le commerce international et l'économie publique pour, entre autres, mesurer les effets d'une fusion d'entreprise, de l'introduction d'un nouveau produit sur le marché ou d'une nouvelle régulation. Du fait de leur simplicité d'estimation, ces modèles devraient intéresser les chercheurs ainsi que les praticiens *antitrust* des cabinets de conseil

[1] À ma connaissance, les seuls modèles d'utilité aléatoire additifs ayant une fonction de demande inverse analytique sont le modèles logit et logit emboîté.

et des autorités de concurrence qui souhaitent éviter des procédures d'estimation complexes et/ou qui sont pressés par le temps.

Plus spécifiquement, cette thèse développe et étudie des modèles de demande inverse pour $J+1$ produits différenciés $j = 0, \ldots, J$ de la forme

$$\sigma_j(\mathbf{s})^{-1} = \ln G_j(\mathbf{s}) + c, \quad j = 0, \ldots, J,$$

où $\mathbf{s}$ est un vecteur de parts de marché, $\ln G_j$ est une fonction dont les propriétés restent à définir et c est une constante commune aux différents produits.

Le premier chapitre de cette thèse construit la classe des modèles *generalized inverse logit* (GIL), lesquels sont des modèles de demande inverse de la forme décrite par l'Equation ci-dessus où $\ln \mathbf{G} \equiv (\ln G_0, \ldots, \ln G_J)$ présente des propriétés spécifiques: $\ln \mathbf{G}$ est telle que $\mathbf{G}$ est homogène de degré un et sa matrice jacobienne est définie positive et symétrique.[2] Ce chapitre montre que chaque modèle de cette classe est cohérent avec un modèle de consommateur représentatif et inclut une grande majorité de modèles d'utilité aléatoire additifs. Il fournit également des méthodes générales pour construire des modèles GIL. Une des méthodes développe des modèles basés sur la construction de nids (i.e., de groupes de produits), lesquels sont analogues à des modèles d'utilité aléatoire additifs qui ont été utilisés à des fins d'estimation de la demande (e.g., le modèle logit ordonné de Small (1987) ou le modèle modèle logit emboîté croisé de Vovsha (1997)). En particulier, il développe le modèle *inverse product differentiation logit* (IPDL), lequel, de manière analogue au modèle de Bresnahan *et al.* (1997), généralise les modèles logit emboîtés, permettant ainsi de capter de façon plus flexible les substitutions entre les produits, y compris de la complémentarité. Cette construction présente toutefois deux limites, lesquelles feront l'objet d'une extension dans le deuxième chapitre. D'abord, elle demande au modélisateur de choisir la structure des nids avant l'estimation. Ensuite, elle implique que la substitution entre produits ne dépend pas directement des caractéristiques des produits – sauf éventuellement celles utilisées pour la construction des nids.

Le second chapitre développe le modèle *flexible inverse logit* (FIL), lequel est un modèle GIL qui dépasse les deux limites associées aux modèles basés sur la construction de nids. Le modèle FIL utilise une structure de nids flexible avec un nid pour chaque pair de produits et un paramètre de nid associé (voir Chu,

[2] Une fonction f de E dans F est dite homogène de degré un si pour tout $\mathbf{x} \in E$, pour tout $\lambda > 0$, $f(\lambda \mathbf{x}) = \lambda f(\mathbf{x})$. La matrice jacobienne est la matrice des dérivées partielles du premier ordre d'une fonction vectorielle.

1989; Koppelman & Wen, 2000; Davis & Schiraldi, 2014) ; il est cohérent avec un modèle appartenant à la classe de modèles de consommateurs hétérogènes maximisateur d'utilité étudiée par Allen & Rehbeck (2019). Les paramètres de nid du modèle FIL sont ensuite projetés dans l'espace des caractéristiques. Basé sur Pinkse *et al.* (2002), ces paramètres sont remplacés par une fonction de la distance entre les produits dans l'espace des caractéristiques. Cette projection permet d'obtenir une substitution entre les produits qui dépend directement des caractéristiques des produits, comme c'est le cas du modèle logit à coefficients aléatoires. La projection utilise également les polynômes de Bernstein afin que la manière dont les substitutions dépendent des caractéristiques soit estimée à partir des données et non postulée. Enfin, des simulations de Monte Carlo ont été menés pour mesurer la capacité du modèle FIL à répliquer les élasticités-prix de la demande de modèles logit à coefficients aléatoires pour des spécifications de l'utilité répandues (absence d'effet revenu, utilité linéaire en le prix, un coefficient aléatoire normalement distribué, etc.). Les résultats des simulations montrent la capacité du modèle FIL à produire des substitutions flexibles.

Le troisième chapitre étudie la micro-fondation du modèle GIL développé dans le premier chapitre de cette thèse. Il montre que les restrictions que le modèle GIL impose sur la fonction de demande inverse sont des conditions nécessaires et suffisantes de cohérence avec un modèle de consommateurs hétérogènes maximisateur d'utilité, connu sous le nom de *perturbed utility model* (PUM) et étudié, entre autres, par Allen & Rehbeck (2019). La preuve de ce résultat implique deux résultats intermédiaires pouvant être considérés comme intéressants en soi. Tout d'abord, tout PUM génère une fonction de demande qui satisfait une légère variante des conditions de Daly-Zachary (voir Daly & Zachary, 1979), laquelle permet de combiner substituabilité et complémentarité en demande. Ensuite, toute fonction de demande satisfaisant ces conditions a une fonction de demande inverse qui est un modèle GIL. Ainsi, par relation d'équivalence, il est montré que les modèles GIL, les PUM et les modèles de demande satisfaisant la variante des conditions de Daly-Zachary fournissent trois modélisations équivalentes du comportement des consommateurs.

Bibliographie

Roy Allen & John Rehbeck. Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054, 2019.

Steven Berry, James Levinsohn, & Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63:841–890, 1995.

Steven Berry, James Levinsohn, & Ariel Pakes. Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review*, 89(3):400–430, 1999.

Steven T Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25:242–262, 1994.

Timothy F Bresnahan, Scott Stern, & Manuel Trajtenberg. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, pages S17–S44, 1997.

C Chu. A paired combinational logit model for travel demand analysis. In *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, pages 295–309. Vol. 4, Western Periodicals, Ventura, CA, 1989.

A Daly & S. Zachary. Improved multiple choice models. In *Determinants of Travel Choice*, pages 337–357. ed. by D. Hensher and Q. Dalvi (eds), London: Teakfield, 1979.

Peter Davis & Pasquale Schiraldi. The flexible coefficient multinomial logit (fc-mnl) model of demand for differentiated products. *The RAND Journal of Economics*, 45(1):32–63, 2014.

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

Pinelopi Koujianou Goldberg. Product differentiation and oligopoly in international markets: The case of the us automobile industry. *Econometrica*, pages 891–951, 1995.

Rachel Griffith, Martin O'Connell, & Kate Smith. Tax design in the alcohol market. *Journal of Public Economics*, 172:20–35, 2019.

Christopher R Knittel & Konstantinos Metaxoglou. Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics*, 96(1):34–59, 2014.

Frank S Koppelman & Chieh-Hua Wen. The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological*, 34(2):75–89, 2000.

Aviv Nevo. Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, 31(3):395–421, 2000.

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

Amil Petrin. Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4):705–729, 2002.

Joris Pinkse, Margaret E Slade, & Craig Brett. Spatial price competition: a semiparametric approach. *Econometrica*, 70(3):1111–1153, 2002.

Kenneth A Small. A discrete choice model for ordered alternatives. *Econometrica*, pages 409–424, 1987.

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

Peter Vovsha. Application of cross-nested logit model to mode choice in tel aviv, israel, metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board*, (1607):6–15, 1997.

Contents

Notice

The three chapters of this dissertation are self-contained research articles. This explains why some information are redundant, and that the term paper is used instead of "chapter". The first chapter of this dissertation is co-authored with Mogens Fosgerau (University of Copenhagen) and André de Palma (CREST, ENS Paris-Saclay, University of Paris-Saclay). Data used in this dissertation, known as the Dominick's Database, are made freely available by the James M. Kilts Center, University of Chicago Booth School of Business.

Chapter 1

The Inverse Product Differentiation Logit Model

1 Introduction

Estimating the demand for differentiated products is of great empirical relevance in industrial organization and other fields of economics. It is important for understanding consumer behavior and for analyzing major economic issues such as the effects of mergers and changes in regulation. Ideally, one would like to employ a model that accommodates rich patterns of substitution, while requiring just regression for estimation.

This paper proposes the Inverse Product Differentiation Logit (IPDL) model, which generalizes the nested logit model by allowing richer patterns of substitution and in particular complementarity (i.e., a negative cross-price elasticity of demand), while being estimable by linear instrumental variables regression.

The IPDL model is relevant for estimating demands for differentiated products that are segmented along multiple dimensions. It generalizes the nested logit models by allowing the segmentation to be non-hierarchical, which is often desirable in applications. At the same time, it maintains the important advantages of the nested logit model. First, its inverse demand has closed form such that numerical inversion of demand is not required. Second, it can be estimated by two-stage least squares regression of market shares on product characteristics and shares related to product segmentation. Third, it is consistent with utility maximization. The IDPL model may therefore be an attractive option in the many empirical applications where the nested logit model would otherwise be used.

The current practice of the demand estimation literature with aggregate data is to assume an additive random utility model (ARUM) (McFadden, 1974) and to estimate it using Berry (1994)'s method to deal with endogeneity of prices and market shares. The logit model is the simplest option, but exhibits the Independence of Irrelevant alternatives (IIA) property. This implies that an improvement in one product draws demand proportionately from all the other products and makes cross-price elasticities independent of how close products are in characteristics space, which is unreasonable in most applications.

The nested logit model with two or more levels generalizes the logit model (see Goldberg, 1995; Verboven, 1996a). This model is commonly used to estimate aggregate demand for differentiated products; some recent examples are Björnerstedt and Verboven (2016) and Berry et al. (2016). The nested logit model has closed-form inverse demand and is conveniently estimated by two-stage least squares. It imposes, however, the restriction that the segmentation of products, i.e., the nesting structure, must be hierarchical, meaning. that each nest on a lower level must be contained within exactly one nest on a higher level. This severely constrains the substitution patterns that the nested logit model can accommodate, since the IIA property still holds within nests and at the nest level. Furthermore, the sequence of segmentation dimensions in the hierarchy is not unique and often not obvious.[1]

The logit and nested logit models belong to the wider class of Generalized Extreme Value (GEV) models developed by McFadden (1978).[2] A number of recent papers have proposed members from this class in order to obtain models with richer substitution patterns. The product differentiation logit model of Bresnahan et al. (1997) extends the nested logit model by allowing the grouping of products to be non-hierarchical. The ordered logit model of Small (1987) and the ordered nested logit model of Grigolon (2018) describe markets having a natural ordering of products.[3] The seminal paper by Berry et al. (1995) overcomes

[1] Hellerstein (2008) writes, concerning the beers market, "[D]emand models such as the multistage budgeting model or the nested logit model do not fit this market particularly well. It is difficult to define clear nests or stages in beer consumption because of the high cross-price elasticities between domestic light beers and foreign light and regular beers. When a consumer chooses to drink a light beer that also is an import, it is not clear if he categorized beers primarily as domestic or imported and secondarily as light or regular, or vice versa."

[2] GEV models are ARUM in which the random utilities have a multivariate extreme value distribution (Fosgerau et al., 2013).

[3] Other papers provide generalizations of the logit model by using semiparametric or nonpara-

the limitations of the nested logit model by specifying a random coefficient logit model, which breaks IIA at the population level. However, the inverse demands of these more general models do not have closed form.

The richer substitution patterns of these models is obtained at the cost of more complex and time-consuming nonlinear estimation procedures such as the nested fixed point (NFP) approach of Berry et al. (1995) or the Mathematical Program with Equilibrium Constraints (MPEC) approach of Dubé et al. (2012), which are associated with issues of local optima and choice of starting values (see e.g., Knittel and Metaxoglou, 2014).

In this paper, we depart from the standard practice by specifying a model in terms of the inverse demand. Given linear-in-parameters utility indexes, the model can then be directly estimated by linear regression using Berry (1994)'s method. More specifically, we propose the IPDL model for products that are segmented along multiple dimensions. The IPDL model extends the nested logit model by allowing arbitrary, non-hierarchical grouping structures (i.e., any partitioning of the choice set in each dimension). It improves on the nested logit model by allowing for richer patterns of substitution and, as we show, even complementarity. This improvement is achieved by removing the constraint that the segmentation should be hierarchical, and it is therefore costless. While the IPDL model requires modelers to define the segmentation, the relative importance of segmentation dimensions can be estimated.

Another important approach in demand estimation is the flexible functional form approach (e.g., the AIDS model of Deaton and Muellbauer, 1980), where the error term has no immediate structural interpretation. By contrast, in this paper, the error term has the structural interpretation of Berry (1994) that it represents product/market-level characteristics unobserved by the modeller but observed by consumers and firms.

The IPDL model belongs to a wider class of inverse demand models, that we label Generalized Inverse Logit (GIL) models. We show that any GIL model is consistent with a representative consumer model (RCM) in which a utility-maximizing representative consumer chooses a vector of nonzero demands, trading off variety against quality. We also show that any ARUM is equivalent to some GIL model. However, the converse is not true, since some GIL models exhibit complementarity, which cannot occur in an ARUM. We establish a new demand inversion result, which extends Berry (1994) and Berry et al. (2013) by

metric methods, see Davis and Schiraldi (2014) for more details.

allowing complementarity. It is often desirable to allow complementarity as important economic questions hinge on the extent to which products are substitutes or complements. In particular, this directly affects the incentives to introduce a new product on the market, to bundle, to merge, etc.[4]

The paper is organized as follows. Section 2 sets the context, introducing the role of demand inversion with the inverse demand of the logit and nested logit models as examples. Section 3 introduces the IPDL model as a generalization of the inverse demand of the nested logit model and shows how to estimate it with aggregate data. Section 4 applies the IPDL model to estimate the demand for ready-to-eat cereals in Chicago, finding that complementarity is pervasive in this market. Section 5 introduces the wider class of GIL models. Section 6 studies its linkages with the ARUM and RCM. Section 7 concludes. A supplement provides Monte Carlo evidence on the IPDL model as well as general methods and examples for building GIL models that go beyond the IPDL model.

2 Motivation

2.1 General Setting: the Role of Demand Inversion

Consider a population of consumers choosing from a choice set of $J+1$ differentiated products, denoted by $\mathcal{J} = \{0, 1, \ldots, J\}$, where products $j = 1, \ldots, J$ are the inside products and product $j = 0$ is the outside good. We consider aggregate data on market shares $s_{jt} > 0$, prices $p_{jt} \in \mathbb{R}$ and K product/market characteristics $\mathbf{x}_{jt} \in \mathbb{R}^K$ for each inside product $j = 1, \ldots, J$ in each market $t = 1, \ldots, T$ (Berry, 1994; Berry et al., 1995; Nevo, 2001). For each market t, the market shares s_{jt} are positive and sum to 1, i.e., $\mathbf{s}_t = \left(s_{0t}, \ldots, s_{Jt}\right) \in \operatorname{int}(\Delta)$, where $\operatorname{int}(\Delta)$ is the interior of the unit simplex in $\mathbb{R}^{J+1}$.

Based on Berry and Haile (2014), let $\delta_{jt} \in \mathbb{R}$ be an index given by

$$\delta_{jt} = \delta\left(p_{jt}, \mathbf{x}_{jt}, \xi_{jt}; \boldsymbol{\theta}_1\right), \quad j \in \mathcal{J}, \quad t = 1, \ldots, T,$$

where $\xi_{jt} \in \mathbb{R}$ is the jt-product/market unobserved characteristics term and $\boldsymbol{\theta}_1$ is a vector of parameters.

[4]See Gentzkow (2007), Ershov et al. (2018), and Iaria and Wang (2019) who investigate these issues empirically.

Consider the system of demand equations

$$\mathbf{s}_t = \boldsymbol{\sigma}(\boldsymbol{\delta}_t;\boldsymbol{\theta}_2), \quad t = 1,\ldots,T, \tag{1}$$

which relates the vector of observed market shares, $\mathbf{s}_t$, to the vector of product indexes in market t, $\boldsymbol{\delta}_t = (\delta_{0t},\ldots,\delta_{Jt})$, through the model, $\boldsymbol{\sigma} = \left(\sigma_0,\ldots,\sigma_J\right)$, where $\boldsymbol{\theta}_2$ is a vector of parameters and

$$\boldsymbol{\sigma}(\cdot;\boldsymbol{\theta}_2) : \mathcal{D} \to \mathrm{int}(\Delta)$$

is an invertible function, with domain $\mathcal{D} \subset \mathbb{R}^{J+1}$.[5]

The market share of the outside good is determined by the identity

$$\sigma_0(\boldsymbol{\delta}_t;\boldsymbol{\theta}_2) = 1 - \sum_{k=1}^{J} \sigma_k(\boldsymbol{\delta}_t;\boldsymbol{\theta}_2), \quad t = 1,\ldots,T.$$

We normalize the index of the outside good, setting $\delta_{0t} = 0$ in each market $t = 1,\ldots,T$.

Several remarks regarding the demand system (1) are in order. First, the unobserved characteristics terms ξ_{jt} are scalars. Second, there is no income effect, since $\boldsymbol{\sigma}$ does not depend on income, and income is implicitly assumed to be sufficiently high that $y > \max_{j\in\mathcal{J}} p_j$. Last, prices p_{jt} and characteristics $\mathbf{x}_{jt}$ enter only through the indexes (in particular, we rule out random coefficients on prices and product characteristics).

Since the function $\boldsymbol{\sigma}$ in Equation (1) is invertible in $\boldsymbol{\delta}_t$, then the inverse demand, denoted by σ_j^{-1}, maps from market shares $\mathbf{s}_t$ to each index δ_{jt} with

$$\delta_{jt} = \sigma_j^{-1}(\mathbf{s}_t;\boldsymbol{\theta}_2), \quad j \in \mathcal{J}, \quad t = 1,\ldots,T. \tag{2}$$

For simplicity, we assume a linear index,

$$\delta_{jt} = \mathbf{x}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \xi_{jt}, \quad j \in \mathcal{J}, \quad t = 1,\ldots,T.$$

Then the unobserved product characteristics terms, ξ_{jt}, can be written as a

[5] Restricting the domain of σ to $\mathcal{D}$ enables the model to be normalized. E.g., $\mathcal{D} = \{\boldsymbol{\delta}_t \in \mathbb{R}^{J+1} : \delta_{0t} = 0\}$ or $\mathcal{D} = \left\{\boldsymbol{\delta}_t \in \mathbb{R}^{J+1} : \sum_{j\in\mathcal{J}} \delta_{jt} = 0\right\}$.

function of the data and parameters $\boldsymbol{\theta}_1 = (\alpha, \boldsymbol{\beta})$ and $\boldsymbol{\theta}_2$ to be estimated,

$$\xi_{jt} = \sigma_j^{-1}(\mathbf{s}_t; \boldsymbol{\theta}_2) + \alpha p_{jt} - \mathbf{x}_{jt}\boldsymbol{\beta}, \quad j \in \mathcal{J}, \quad t = 1, \ldots, T. \tag{3}$$

The unobserved product characteristics terms ξ_{jt} represent the structural error terms of the model, since we assume that they are observed by consumers and firms but not by the modeller. In addition, prices and market shares in the right-hand side of Equation (3) are endogenous, i.e., they are correlated with the structural error terms ξ_{jt}.[6] Then, following Berry (1994), we can estimate demands (1) based on the conditional moment restrictions

$$\mathbb{E}\left[\xi_{jt} | \mathbf{z}_t\right] = 0, \quad j \in \mathcal{J}, \quad t = 1, \ldots, T,$$

provided that there exists appropriate instruments $\mathbf{z}_t$ for the endogenous prices and market shares.

2.2 Closed-Form and Linear-in-Parameters Inverse Demands

Since the seminal papers by Berry (1994) and Berry et al. (1995), the standard practice of the demand estimation literature with aggregate data has been to specify an ARUM and to compute the corresponding demands, which then must be inverted numerically during estimation.[7] In this paper, we instead directly specify inverse demands of the form

$$\sigma_j^{-1}(\mathbf{s}_t; \boldsymbol{\theta}_2) = \ln G_j(\mathbf{s}_t; \boldsymbol{\theta}_2) + c_t, \quad j \in \mathcal{J}, \tag{4}$$

where the vector function $\mathbf{G} = (G_0, \ldots, G_J)$ is invertible as a function of $\mathbf{s}_t \in \operatorname{int}(\Delta)$, and where $c_t \in \mathbb{R}$ is a market-specific constant.[8] Combining with Equation (2), this amounts to

$$\ln G_j(\mathbf{s}_t; \boldsymbol{\theta}_2) = \delta_{jt} - c_t. \tag{5}$$

[6] Prices are likely to be endogenous since firms may consider both observed and unobserved product characteristics when setting prices. Market shares are endogenous by construction since they are defined by the system of equations (1), where each demand depends on the entire vectors of endogenous prices and unobserved product characteristics.

[7] To our knowledge, the logit and the nested logit models are the only ARUM that yield closed-form inverse demands.

[8] Compiani (2019) adopts a similar approach, but proposing to nonparametrically estimate inverse demands for differentiated products based on aggregate data.

When $\ln G_j$ is linear in parameters $\boldsymbol{\theta}_2$, estimation amounts to linear regression, which makes two-stage least squares (2SLS) easily applicable and (empirical) identification clear.

The logit and the nested logit models have closed-form and linear-in-parameters inverse demands that satisfy Equation (4). For the logit model,

$$\ln G_j(\mathbf{s}_t) = \ln\left(s_{jt}\right), \quad j \in \mathcal{J},$$

so that its inverse demand its given by the following well-known expression (Berry, 1994)

$$\sigma_j^{-1}(\mathbf{s}_t) = \ln\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt}.$$

For the two-level nested logit model, which partitions the choice set into groups,

$$\ln G_j(\mathbf{s}_t;\mu) = (1-\mu)\ln\left(s_{jt}\right) + \mu \ln\left(\sum_{k\in\mathcal{G}(j)} s_{kt}\right), \quad j \in \mathcal{J},$$

where $\mathcal{G}(j)$ is the set of products grouped with product j and $\mu \in (0,1)$ is the nesting parameter (see Berry, 1994).

For the three-level nested logit model, which extends the two-level nested logit model by further partitioning groups into subgroups,

$$\ln G_j(\mathbf{s}_t;\mu_1,\mu_2) = \left(1 - \sum_{d=1}^{2}\mu_d\right)\ln\left(s_{jt}\right) + \mu_1 \ln\left(\sum_{k\in\mathcal{G}_1(j)} s_{kt}\right) + \mu_2 \ln\left(\sum_{k\in\mathcal{G}_2(j)} s_{kt}\right),$$

where the parameters satisfy $\sum_{d=1}^{2}\mu_d < 1$, $\mu_d \geq 0$, $d = 1,2$, and where $\mathcal{G}_1(j)$ and $\mathcal{G}_2(j)$ are the sets of products belonging the same group and to the same subgroup as product j, respectively.[9]

The logit and nested logit models have the important advantage that they boil down to linear regression models (Berry, 1994). For example, for the logit model,

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \xi_{jt}, \quad j = 1,\ldots,J, \quad t = 1,\ldots,T.$$

The logit model requires just one instrument for price and the two-level nested

[9] Indeed, setting $\gamma_1 = \mu_1 + \mu_2$ and $\gamma_2 = \mu_1$, we recover Equation (10) of Verboven (1996a) and the model satisfies the constraint $0 \leq \gamma_2 \leq \gamma_1 < 1$ that makes it consistent with random utility maximization.

logit model requires one instrument for price and one for the endogenous shares. As a consequence, both models allow very large choice sets involving thousands of products. However, the logit and nested logit models impose strong restrictions on the substitution patterns that can be accommodated.

In the next section, we introduce the inverse product differentiation logit (IPDL) model, which extends the inverse demand of the nested logit model in the same way that the product differentiation logit model of Bresnahan et al. (1997) extends the nested logit model; we take the IPDL model to data on ready-to-eat cereals in Section 4.

3 The Inverse Product Differentiation Logit (IPDL)

Specification of the Model Suppose that each market exhibits product segmentation along D *dimensions*, indexed by d. Each dimension d defines a finite number of *groups* of products, such that each product belongs to exactly one group in each dimension. The grouping structure is exogenous and, for simplicity, assumed to be common across markets.

Let $\boldsymbol{\theta}_2 = (\mu_1, \ldots, \mu_D)$, where $\sum_{d=1}^{D} \mu_d < 1$ and $\mu_d \geq 0$, $d = 1, \ldots, D$, and let $\mathcal{G}_d(j)$ be the set of products grouped with product j on dimension d. The IPDL model has inverse demands that are given by Equation (4), where $\ln G_j$ is defined as

$$\ln G_j(\mathbf{s}_t; \boldsymbol{\theta}_2) = \left(1 - \sum_{d=1}^{D} \mu_d\right) \ln\left(s_{jt}\right) + \sum_{d=1}^{D} \mu_d \ln\left(\sum_{k \in \mathcal{G}_d(j)} s_{kt}\right). \tag{6}$$

We show below that inverse demands (6) are invertible, such that it is possible to compute the IPDL demands.[10] We show in Section 6 that the IDPL demand is consistent with utility maximization.

We say that two products are of the same *type* if they belong to the same group on all dimensions. We assume that the outside good is the only product of its type, which implies that

$$\ln G_0(\mathbf{s}_t; \boldsymbol{\theta}_2) = \ln(s_{0t}) = \delta_{0t} - c_t = -c_t. \tag{7}$$

[10] Invertibility of $\ln \mathbf{G} = \left(\ln G_0, \ldots, \ln G_J\right)$ is shown using Proposition 1. The key assumption that ensures invertibility is that $\sum_{d=1}^{D} \mu_d < 1$, which means that a positive weight is assigned to the terms $\ln(s_{jt})$.

The IPDL model extends the nested logit model by allowing arbitrary, non-hierarchical grouping structures, i.e., any partitioning of the choice set in each dimension. Figure 1 illustrates the competing hierarchical and non-hierarchical grouping structures used for the empirical application in Section 4. The freedom in defining grouping structures can be used to build inverse demand models that are similar in spirit to GEV models based on nesting, which have proved useful for demand estimation purposes (Train, 2009, Chap. 4). For example, like Small (1987) and Grigolon (2018), it is possible to define grouping structures that describe markets where products are naturally ordered.

It is important to note that the parametrization of the IPDL model does not depend on the number of products, but instead on the number of segmentation dimensions. This is important because it implies that the IPDL model can handle markets involving very many products.

Estimation of the IPDL Model Combining Equations (6) and (7), the IPDL model boils down to a linear regression model of market shares on product characteristics and share terms

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}\beta - \alpha p_{jt} + \sum_{d=1}^{D} \mu_d \ln\left(\frac{s_{jt}}{\sum_{k\in\mathcal{G}_d(j)} s_{kt}}\right) + \xi_{jt}, \tag{8}$$

for all inside products $j = 1,\ldots,J$ in each market $t = 1,\ldots,T$.

Equation (8) has the same form as the logit and nested logit equations (see Berry, 1994; Verboven, 1996a), except for the share terms. Under the standard assumption that product characteristics $\mathbf{x}_{jt}$ are exogenous, we must therefore find at least $D+1$ instruments: one instrument for price and for each of the D share terms.

Following the prevailing literature (see e.g., Berry and Haile, 2014; Reynaert and Verboven, 2014; Armstrong, 2016), both cost shifters and BLP instruments are good candidates for instruments. Cost shifters are appropriate instruments for prices but may not be appropriate for market shares because costs affect the market shares only through prices. BLP instruments, which are functions of the characteristics of competing products and are commonly used to instrument prices, are also useful to instrument market shares. In theory, BLP instruments generally suffice for identification.[11] However, in practice they may be weak and

[11] See Armstrong (2016) for a discussion on the validity of BLP instruments as the number of

then cost shifters are required.

Following Verboven (1996a) and Bresnahan et al. (1997), the BLP instruments of the IPDL model include, for each dimension, the sums of characteristics for other products of the same group as well as the number of products within each group. These instruments reflect the degree of substitution and the closeness of products within a group and are therefore likely to affect prices and within-group market shares. The same instruments can also be computed for products of the same type. Lastly, we can exploit the ownership structure of the market and compute the same instruments while distinguishing products of the same firms from products of competing firms. The idea is that prices, and thus market shares, depend on the ownership structure since multi-product firms set prices so as to maximize their total profits.

Links to Discrete Choice Models We show below that the IPDL model is consistent with a representative consumer model (RCM) with taste for variety and without income effect. The RCM assumes the existence of a variety-seeking representative consumer who aggregates a population of consumers and chooses some quantity of every product, trading off variety against quality. It has been a workhorse of the international trade literature since Dixit and Stiglitz (1977) and Krugman (1979), and has also been used for demand estimation purposes (e.g., Pinkse and Slade, 2004).

Specifically, as shown below, the IPDL model is consistent with a representative consumer, endowed with income y, who chooses a vector $\mathbf{s}_t \in \operatorname{int}(\Delta)$ of nonzero market shares in market t so as to maximize the utility

$$\alpha y + \sum_{j \in \mathcal{J}} \delta_{jt} s_{jt} - \sum_{j \in \mathcal{J}} s_{jt} \ln G_j(\mathbf{s}_t), \tag{9}$$

where G_j is defined by Equations (6) and (7), and where δ_j is a linear-in-price index. The second term in Equation (9) captures the net utility derived from the consumption of $\mathbf{s}_t$ in the absence of interaction among products and the last term expresses taste for variety.

As mentioned above, the IPDL model has the nested logit model, and thus the logit model, as special cases: the logit is obtained when product segmentation does not matter, and the nested logit model is obtained when the grouping

products increases.

structure is hierarchical. Thus some IDPL models are ARUM. On the other hand, as shown below and in contrast to any ARUM, some IDPL models allow complementarity.[12]

Complementarity We use the standard definition of complementarity and substitutability in the absence of income effect (Samuelson, 1974), defining complementarity (resp., substitutability) as a negative (resp., positive) cross-price derivative of demand.[13] Proposition 4 in Appendix A.3 provides some properties of the IPDL model regarding the patterns of substitution, including the matrix of price derivatives of demand.

The IPDL model allows complementarity. To see this, suppose there are 3 inside products and one outside good. Inside products are grouped according to two dimensions: for the first dimension, product 1 is in one group, and products 2 and 3 are in a second group; for the second dimension, products 1 and 2 are in one group, and product 3 is in a second group. Products 1 and 3 are complements if the derivative of the demand for product 3 with respect to the price of product 1 is negative, that is, if[14]

$$(1-\mu_1-\mu_2)(s_1+s_2)(s_2+s_3)-\mu_1\mu_2 s_0 s_2 > 0,$$

which simplifies to $4/3 > \mu_1\mu_2/(1-\mu_1-\mu_2)$ for $s_0 = 1/2$ and $s_1 = s_2 = s_3 = 1/6$. In particular, products 1 and 3 are complements for $\mu_1 = \mu_2 = 1/3$, but are substitutes for $\mu_1 = \mu_2 = 0.45$. With the representative consumer interpretation, the parameter $\mu_0 = 1-\sum_{d=1}^{D}\mu_d$ measures taste for variety over all products of the choice set and each parameter μ_d, for $d \geq 1$, measures taste for variety across groups of products according to dimension d: higher μ_d means that variety at the level of groups of products matters more, meaning that products in the same group in dimension d are more similar (see Verboven, 1996b, for a similar interpretation of the nesting parameter of the nested logit model). Then complementarity in the IPDL model arises in a very intuitive way, due to taste for variety at the group level.

In Section 1 of the supplement, we provide some simulation results investi-

[12] It would be of interest to establish conditions under which the IDPL model is equivalent to an ARUM.

[13] This definition is different from the one used by Gentzkow (2007) in the context of an ARUM defined over bundles of products.

[14] See Proposition 4 in Appendix A.3.

gating the patterns of substitution. We find that: (i) products of the same type are always substitutes, while products of different types may be substitutes or complements; and (ii) closer products into the characteristics space used to form product types (i.e., higher values of μ_d and/or whether products belong to the same groups or not) have higher cross-price elasticities.

4 Empirical Illustration: Demand for Cereals

In this section, we illustrate the IPDL model by estimating the demand for ready-to-eat (RTE) cereals in Chicago in 1991 – 1992. This market has been studied extensively (see e.g., Nevo, 2001; Michel and Weiergraeber, 2019) and it is known to exhibit product segmentation. We compare the results (in terms of elasticities and goodness-of-fit) from the IPDL model to those from two alternative nested logit models.

4.1 Data

Databases We use store-level scanner data from the Dominick's Database, made available by the James M. Kilts Center, University of Chicago Booth School of Business. We supplement with data on the nutrient content of the RTE cereals (sugar, energy, fiber, lipid, sodium, and protein) from the USDA Nutrient Database for Standard Reference and with monthly sugar prices from the website www.indexmundi.com.

For our analysis, we use data from 83 Dominick's stores and focus on the 50 largest products in terms of sales (e.g., Kellogg's Special K), where a product is a cereal (e.g., Special K) associated to its brand (e.g., Kellogg's). We define a market as a store-month pair. Prices of a serving (i.e., 35 grammes) and market shares are computed following Nevo (2001). See Appendix B for more details on the data.

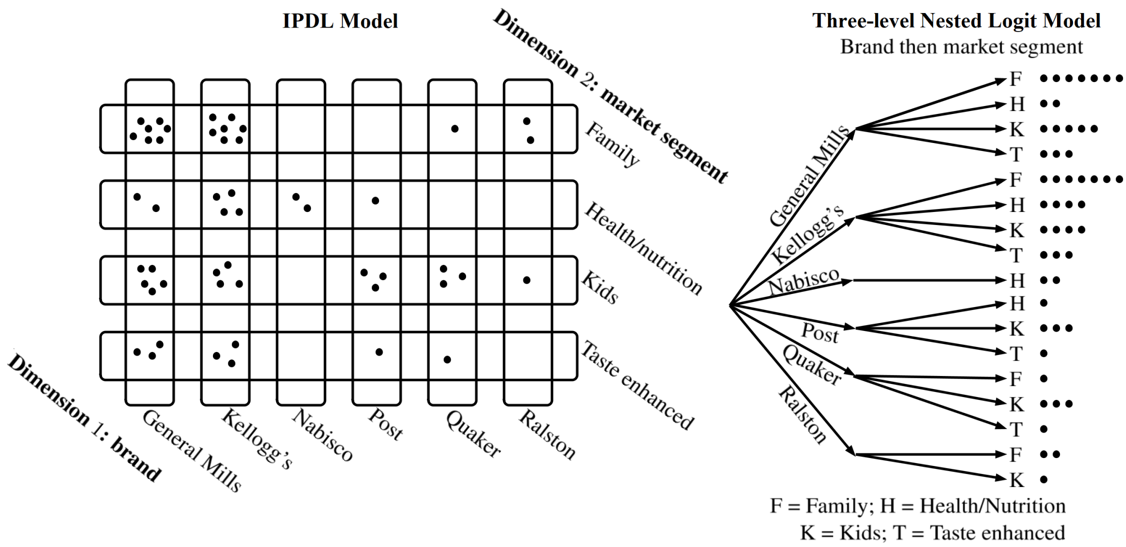
Product Segmentation Based on the observations below, we consider two segmentation dimensions. The first dimension is brand, where the brands are General Mills, Kellogg's, Nabisco, Post, Quaker, and Ralston. The second dimension is market segment, where the market segments are family, kids, health/nutrition, and taste enhanced (see e.g., Nevo, 2001). These two dimensions are combined to form 17 product types among the 50 products.

Brands play an important role: Kellogg's is the largest company and has large market shares in all market segments; and General Mills, the second largest company, is especially popular in the family and kids segments. Taken together, Kellogg's and General Mills account for around 80 percent of the market. As regards market segments, the family and kids segments dominate and account for almost 70 percent of the market.

Table 1 shows the average nutrient content of the cereals grouped by brand and market segment. As expected, cereals for health/nutrition contain less sugar, more fiber, less lipid, and less sodium, and are less caloric. By contrast, cereals for kids contain more sugar and more calories. Moreover, Nabisco offers cereals with less sugar and less calories, while Quaker and Ralston offer cereals with more calories. The two dimensions therefore proxy, at least partially, for the nutrient content of the cereals.

Figure 1 illustrates the grouping structure of the IPDL model (left panel) and of the three-level nested logit model where products are grouped first by brand and then by market segment (right panel).

Figure 1: Product Segmentation on the Cereals Market



Each dot illustrates the location of a cereal in the grouping structure.

Table 1: Average by Market Segment and by Brand

Dimensions	Sugar g/serve	Energy kcal/serve	Fiber g/serve	Lipid g/serve	Sodium mg/serve	Protein g/serve	N
Brand (dimension 1)							
General Mills	9.92	132.09	1.99	1.51	230.69	2.65	17
	(4.67)	(7.69)	(0.98)	(0.82)	(60.83)	(0.83)	
Kellogg's	9.58	127.50	2.47	0.85	228.49	2.88	18
	(5.52)	(11.16)	(2.81)	(0.96)	(103.93)	(1.43)	
Nabisco	0.25	125.48	3.43	0.58	2.10	3.83	2
	(0.09)	(0.74)	(0)	(0)	(1.98)	(0.02)	
Post	12.02	130.76	2.09	1.03	212.03	2.49	5
	(4.64)	(14.83)	(2.02)	(0.78)	(22.31)	(1.15)	
Quaker	8.50	139.44	2.26	2.43	159.88	3.59	5
	(4.04)	(9.20)	(0.66)	(1.86)	(94.60)	(1.15)	
Ralston	7.09	138.48	0.58	0.51	305.43	2.04	3
	(6.61)	(1.41)	(0.08)	(0.65)	(71.57)	(0.39)	
Market Segment (dimension 2)							
Family	7.54	130.41	2.22	0.99	269.66	2.88	17
	(5.27)	(9.83)	(2.61)	(0.71)	(88.64)	(1.03)	
Health/nutrition	5.03	122.54	3.16	0.54	168.54	3.84	9
	(3.69)	(5.78)	(1.31)	(0.21)	(133.62)	(1.35)	
Kids	13.40	137.75	1.00	1.35	211.38	2.01	16
	(4.17)	(3.80)	(0.69)	(0.79)	(44.77)	(0.87)	
Taste enhanced	9.70	129.28	3.32	2.22	166.43	3.16	8
	(2.05)	(15.50)	(1.12)	(1.93)	(76.38)	(0.34)	
Total	9.31	131.16	2.17	1.22	216.29	2.82	50
	(5.21)	(10.21)	(1.92)	(1.08)	(93.53)	(1.15)	

Notes: Standard deviations are reported in parentheses. Column "N" gives the number of products by market segment and by brand. Averages and standard deviations are computed over products (without taking into account of their market shares).

4.2 Demand Estimation

Specification We estimate Equation (8), where $\mathbf{x}_{jt}$ includes a constant, the nutrients mentioned above and a dummy for promotion. Following Bresnahan et al. (1997), we include fixed effects for brands (ξ_b) and market segments (ξ_s), which capture market-invariant observed and unobserved brand-specific and market segment-specific characteristics, respectively. The advantages provided by the two dimensions are therefore parametrized by the fixed effects ξ_b and ξ_s, which measure the extent to which belonging to a group shifts the demand for the product, as well as the parameters for groups μ_1 and μ_2, which measure the extent to which products within a group are protected from substitution from products in other groups along each dimension. Lastly, we include fixed effects for month (ξ_m), and store (ξ_{st}), which capture monthly unobserved determinants of demand and time-invariant store characteristics, respectively.

The unobserved product characteristics terms are therefore given by

$$\xi_{jt} = \xi_b + \xi_s + \xi_m + \xi_{st} + \tilde{\xi}_{jt},$$

where $\tilde{\xi}_{jt}$, the structural error that remain in ξ_{jt}, capture the unobserved product characteristics varying across products and markets that are not included into the model (e.g., changes in shelf-space, positioning of the products among others), which affect consumers utility and that consumers and firms (but not the modeller) observe so that they are likely to be correlated with prices and market shares.

Instruments We use two sets of instruments. First, as cost-based instruments, we form the price of the cereal's sugar content of a serve (i.e., sugar content in a serve times the sugar monthly price), which we interact with brand fixed effects. Multiplying the price of sugar by the sugar content allows the instrument to vary by product; and interacting this with fixed effects allows the price of sugar to enter the production function of each firm differently.

Second, we form BLP instruments by using other products' promotional activity in a given market, which varies both across stores for a given month and across months for a given store: for a given product, other products' promotional activity should affect consumers' choices, and should thus be correlated with the price and market share of that product, but not with the error term.[15]

[15]We do not use functions of the nutrient content of the cereals as instruments since by construction of the data they are invariant across markets. We treat promotion as an exogenous variable since, at Dominick's Finer Foods, the promotional calendar is known several weeks in advance of the weekly price decisions. One concern about the use of promotions to form instruments is that promotions could be advertised. If it was the case, this would mean that promotions are not exogenous and cannot be used as instruments. However, we do not observe advertising in the data, which is therefore part of the error term, and, in turn, we assume that promotions

Specifically, we use the number of other promoted products of rival firms and the number of other promoted products of the same firm, which we interact with brand name fixed effects. We also use these numbers over products belonging to the same market segment, which we interact with market segment fixed effects.

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

Parameter Estimates Table 2 presents the 2SLS demand estimates from the IPDL and the three-level nested logit models with groups for market segment on top (3NL1) and with groups for brand on top (3NL2), in columns (1), (2), and (3), respectively.

Consider first the results from the IPDL model. As expected, the estimated parameters on the negative of price (α) and on promotion (β) are significantly positive. The estimated parameters for groups have magnitude and sign that conform to the assumptions of the IPDL model, $\mu_1 \geq 0$, $\mu_2 \geq 0$ and $1 - \mu_1 - \mu_2 > 0$; no constraints were imposed on the parameters during the estimation. These estimates imply that there is product segmentation along both dimensions: for cereals of the same market segment, cereals of the same brand are closer substitutes than cereals of different brands; and for cereals of the same brand, cereals within the same market segment are closer substitutes than cereals from different market segments. Overall, cereals of the same type are closer substitutes.

We find that the brand reputation of the cereals confers a significant advantage to products from General Mills and Kellogg's and that cereals for family have a significant advantage. In addition, we find that $\mu_1 > \mu_2$, which means that the market segments confer more protection from substitution than brand reputation does (cereals of the same market segment are more protected from cereals from different market segments than cereals of the same brand are from cereals of different brands).

are not followed by advertising. See Michel and Weiergraeber (2019) who also use promotion to form instruments.

Table 2: Parameter Estimates of Demand

	(1) IPDL		(2) 3NL1		(3) 3NL2	
Price ($-\alpha$)	-1.83	(0.12)	-2.91	(0.12)	-4.10	(0.16)
Promotion (β)	0.088	(0.003)	0.102	(0.003)	0.144	(0.004)
Constant (β_0)	-0.697	(0.059)	-0.379	(0.065)	-0.195	(0.076)
Nesting Parameters (μ)						
Market Segment/Group (μ_1)	0.626	(0.009)	0.771	(0.008)	0.668	(0.011)
Brand/Subgroup (μ_2)	0.232	(0.009)	0.792	(0.007)	0.709	(0.010)
FE Brands (ξ_b)						
Kellogg's	0.024	(0.005)	-0.056	(0.003)	0.104	(0.006)
Nabisco	-0.754	(0.024)	-0.218	(0.011)	-2.11	(0.02)
Post	-0.485	(0.014)	-0.187	(0.008)	-1.36	(0.01)
Quaker	-0.553	(0.015)	-0.329	(0.014)	-1.51	(0.01)
Ralston	-0.732	(0.025)	-0.200	(0.011)	-2.13	(0.02)
FE Market Segments (ξ_s)						
Health/nutrition	-0.672	(0.010)	-0.855	(0.008)	-0.069	(0.005)
Kids	-0.433	(0.009)	-0.529	(0.009)	0.071	(0.005)
Taste enhanced	-0.710	(0.010)	-0.903	(0.007)	-0.088	(0.006)
Observations	99281		99281		99281	
RMSE	0.210		0.242		0.270	

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

Model Selection and Robustness The estimates from the two nested logit models satisfy $\mu_2 > \mu_1$, which means that they are both consistent with random utility maximization. Neither nested logit model can then be rejected on this criterion.

The three estimated models are non-nested and have the same number of estimated parameters. Then the non-nested test of Rivers and Vuong (2002) can be used to determine which best fits the data. We find that the test strongly rejects both nested logit models in favor of the IPDL model.[16]

In many situations, consumers face choices involving a very large number

[16] The test statistic is given by $\sqrt{J \times T}\left(\hat{Q}_1 - \hat{Q}_2\right)/\hat{\sigma}$, where $\hat{Q}_i$ is the value of the 2SLS objective function of model i evaluated at the demand estimates, and $\hat{\sigma}^2$ is the estimated value of the variance of the difference between $\hat{Q}_i$'s. The null hypothesis is that the two non-nested models are asymptotically equivalent; the first (resp., second) alternative hypothesis is that model 1 (resp., model 2) is asymptotically better than model 2 (resp., model 1). This statistic must be evaluated against the standard normal distribution and we estimate $\hat{\sigma}^2$ using 500 bootstrap replications. The test statistics of the two nested logit models (model 1) against the IDPL model (model 2) are 1509.77 and 3644.43, respectively.

of products (e.g., choice of a car or of a RTE cereal). We have estimated an alternative specification in which we define products as cereal-brand-store combinations, as it is commonly done in the vertical relationships literature (see e.g., Villas-Boas, 2007), and markets as months. The resulting specification, which has more than $4,000$ products, leads to very similar parameter estimates, thereby indicating that the results are fairly robust to the definitions of products and markets.

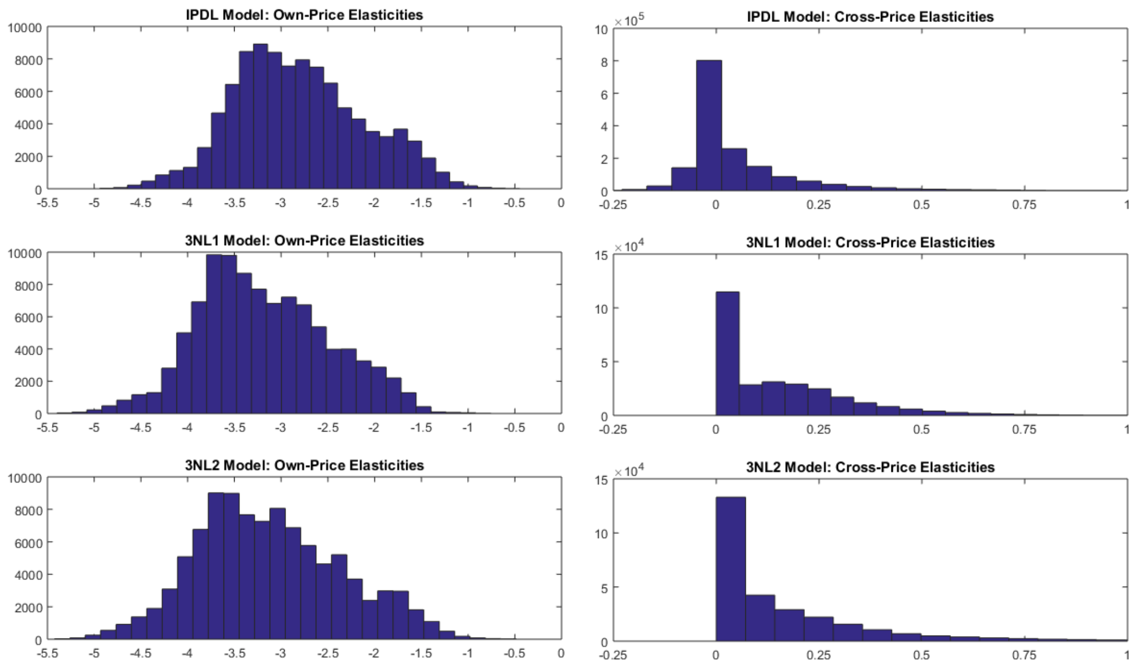
Substitution Patterns. Figure 2 presents the estimated density of the own- and cross-price elasticities of demands of the IPDL and the two nested logit models (see Section 3 of the supplement for the estimated own- and cross-price elasticities of demands, averaged across markets and product types).

The estimated own-price elasticities are in line with the literature (see e.g., Nevo, 2001). On average, the estimated own-price elasticity of demands is -2.815 for the IPDL model, -3.187 for the 3NL1 model and -3.124 for the 3NL2 model. However, there is an important variation in price responsiveness across product types: for the IPDL model, own-price elasticities range from -3.560 for cereals for kids produced by General Mills to -1.388 for cereals for health/nutrition produced by Post; for the 3NL1 model, they range from -3.923 for cereals for kids produced by Ralston to -1.868 for cereals for health/nutrition produced by Post; and for the 3NL2 model, they range from -3.975 for cereals for kids produced by General Mills to -1.488 for cereals for health/nutrition produced by Post.

Consider now the cross-price elasticities. Among the 50×50 different cross-price elasticities that the IPDL model yields, 49.5 percent (resp., 50.5 percent) are negative (resp., positive), meaning about one half of all pairs of cereals are complements. Note that the presence of complementarity is likely to reduce competition in the cereals market compared to a case with no complementarity. Iaria and Wang (2019) also find that complementarity is pervasive in the RTE cereals market. Complementarity can arise for many reasons, including taste for variety and shopping costs.

Products of the same brand are always substitutes, which means that there is cannibalization effect; likewise, products from the same market segment are all substitutes. Products of different brands and of different market segments may or may not be complements.

Figure 2: Estimated Price Elasticities of Demands



5 The Generalized Inverse Logit Model

In this section, we introduce the Generalized Inverse Logit (GIL) class of models, which includes the IPDL model as a special case and hence also the logit and nested logit models. Proofs for this section are provided in Appendix A.4 along with formal statements of the results. To ease exposition, we omit notation for parameters $\boldsymbol{\theta}_2$ and market t.

Definition 1. GIL models are inverse demands of the form (5), i.e.,

$$\ln G_j(\mathbf{s}) = \delta_j - c, \quad j \in \mathcal{J}, \tag{10}$$

where $c \in \mathbb{R}$ is a market-specific constant and $\ln \mathbf{G} = \left(\ln G_0, \ldots, \ln G_J\right)$ is an inverse GIL demand.

An inverse GIL demand is a function $\ln \mathbf{G}$, where $\mathbf{G} : (0,\infty)^{J+1} \to (0,\infty)^{J+1}$ is homogeneous of degree one and where the Jacobian $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is positive definite and symmetric.

This definition immediately implies that the IDPL model is also a GIL model. Section 2 of the supplement provides a range of general methods for building inverse GIL demands along with illustrative examples that go beyond the IPDL model, which is the focus of the paper. As stated in the following proposition, an inverse GIL demand is injective and hence invertible on its range.

Proposition 1. Let $\ln \mathbf{G}$ be an inverse GIL demand. Then $\ln \mathbf{G}$ is injective on $\operatorname{int}(\Delta)$.

We denote the inverse function as $\mathbf{H} = \mathbf{G}^{-1}$. Inverting Equation (10) and using that demands sum to one together with the homogeneity of $\mathbf{G}$ leads to the demand functions

$$s_j = \sigma_j(\boldsymbol{\delta}) = \frac{H_j\left(e^{\boldsymbol{\delta}}\right)}{\sum_{k \in \mathcal{J}} H_k\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J}. \tag{11}$$

This expression generalizes the logit demands in a nontrivial way through the presence of the function $\mathbf{H}$.

In addition, consider any vector of market shares $\mathbf{s} \in \operatorname{int}(\Delta)$. Then, holding $\delta_0 = 0$, the injectivity of the inverse GIL demand ensures that there exists a unique vector of indexes $\boldsymbol{\delta} = \left(0, \delta_1, \ldots, \delta_J\right)$ that rationalizes demand, i.e., $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$.

Using that demands satisfy Roy's identity $\partial CS\left(e^{\boldsymbol{\delta}}\right)/\partial \delta_j = \sigma_j(\boldsymbol{\delta})$, it can easily be shown that the consumer surplus is

$$CS(\boldsymbol{\delta}) = \ln\left(\sum_{k \in \mathcal{J}} H_k\left(e^{\boldsymbol{\delta}}\right)\right),$$

up to an additive constant. Note that the consumer surplus is simply the logarithm of the denominator of the demands in Equation (11), just as in the case of the logit model.

Using that demands sum to one, we obtain that the market-specific constant is equal to the consumer surplus $c = CS(\boldsymbol{\delta})$, which combined with Equation (10), shows that GIL models satisfy

$$\delta_j = \ln G_j(\mathbf{s}) + CS(\boldsymbol{\delta}), \quad j \in \mathcal{J}. \tag{12}$$

Differentiating (12) with respect to $\boldsymbol{\delta}$, we find that the matrix of demand derivatives $\partial \sigma_j / \partial \delta_i$ is given by

$$\mathbf{J}_{\boldsymbol{\sigma}}(\boldsymbol{\delta}) = \left[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})\right]^{-1} - \mathbf{s}\mathbf{s}^{\intercal}, \tag{13}$$

where $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$. Given the absence of income effects, the matrix (13) is the Slutsky matrix. This is symmetric and positive semi-definite, which implies that GIL demands are non-decreasing in their own index δ_j, $\partial\sigma_j/\partial\delta_j \geq 0$.

The class of GIL models accommodates patterns that go beyond those of standard ARUM. In particular, it allows for complementarity: this is for example the case of the IPDL model, which is a member of the class of GIL models. Our invertibility result in Proposition 1 therefore extends Berry (1994)'s invertibility result, which restricts the products to be strict substitutes. Proposition 1 also extends Berry et al. (2013), who show invertibility for demands that satisfy their "connected substitutes" conditions, which rule out complementarity.[17]

6 Relationships between Models

This section puts the GIL and the IPDL models into perspective by showing how they relate to the representative consumer model (RCM) and to the additive random utility model (ARUM).

6.1 Representative Consumer Model

Consider a representative consumer facing the choice set of differentiated products, $\mathcal{J}$, and a homogeneous numéraire good, with demands for the differentiated products summing to one. Let p_j and v_j be the price and the quality of product $j \in \mathcal{J}$, respectively.

The price of the numéraire good is normalized to 1 and the representative consumer's income y is sufficiently high ($y > \max_{j\in\mathcal{J}} p_j$) to guarantee that consumption of the numéraire good is positive.

In this subsection, we show that the inverse GIL demand $\ln \mathbf{G}$ is consistent with a representative consumer who chooses a consumption vector $\mathbf{s} \in \Delta$ of market shares of the differentiated product and a quantity $z \geq 0$ of the numéraire

[17]The connected substitutes structure requires two conditions: (i) products are weak gross substitutes, i.e., everything else equal, an increase in δ_i weakly decreases demand σ_j for all other products; and (ii) the "connected strict substitution" condition holds, i.e., there is sufficient strict substitution between products to treat them in one demand system. In contrast to ours, Berry et al. (2013)'s result does not require that demand $\boldsymbol{\sigma}$ is differentiable. Demand systems with complementarity may be covered by Berry et al. (2013)'s result in cases where a suitable transformation of demand can be found such that the transformed demand satisfies their conditions. They provide no general result on how such a transformation could be found. Our result allows complementarity without requiring such a transformation to be found.

good, so as to maximize her direct utility

$$\alpha z + \sum_{j \in \mathcal{J}} v_j s_j - \sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s}) \tag{14}$$

subject to the budget constraint and the constraint that demands sum to one,

$$\sum_{j \in \mathcal{J}} p_j s_j + z \leq y \quad \text{and} \quad \sum_{j \in \mathcal{J}} s_j = 1, \tag{15}$$

where $\alpha > 0$ is the marginal utility of income. The first two terms of the direct utility (14) describe the utility that the representative consumer derives from the consumption $(\mathbf{s}, z)$ of the differentiated products and the numéraire in the absence of interaction among them. The third term is a strictly concave function of $\mathbf{s}$ that expresses the representative consumer's taste for variety (see Lemma 4 in Appendix A.5).

Let $\delta_j = v_j - \alpha p_j$ be the net utility that the consumer derives from consuming one unit of product $j \in \mathcal{J}$. The utility maximization program (14)-(15) leads to first-order conditions for interior solution

$$\ln G_j(\mathbf{s}) + c = \delta_j, \tag{16}$$

which are of the form of Equation (10) defining the inverse GIL demand.

We state this observation as a proposition and give a detailed proof in Appendix A.5.

Proposition 2. The GIL model (16) is consistent with a representative consumer who maximizes utility (14) subject to constraints (15).

This proposition thus extends Anderson et al. (1988) and Verboven (1996b)'s results that the logit and the nested logit models are consistent with a utility maximizing representative consumer.

6.2 Additive Random Utility Model

We now turn to the Additive Random Utility Model. A population of consumers face the choice set of differentiated products, $\mathcal{J}$, and associate a deterministic utility component $\delta_j = v_j - \alpha p_j$ to each product $j \in \mathcal{J}$. Each individual consumer chooses the product that maximizes her indirect utility given by[18]

$$u_j = \delta_j + \varepsilon_j, \quad j \in \mathcal{J}, \tag{17}$$

where the vector of random utility components $\boldsymbol{\varepsilon} = \left(\varepsilon_0, \ldots, \varepsilon_j, \ldots, \varepsilon_J\right)$ follows a joint distribution with finite means that is absolutely continuous, fully supported on $\mathbb{R}^{J+1}$ and independent of $\boldsymbol{\delta}$. These assumptions are standard in the discrete choice literature. They imply that utility ties occur with probability 0, that the choice probabilities are all everywhere positive, and that random coefficients are ruled out. Specific distributional assumptions for $\boldsymbol{\varepsilon}$ lead to specific models such as the logit model, the nested logit model, the probit model, etc.

The probability that a consumer chooses product j is

$$P_j(\boldsymbol{\delta}) = \Pr\left(u_j \geq u_i, \ \forall i \neq j\right), \quad j \in \mathcal{J}.$$

Let $\overline{CS} : \mathbb{R}^{J+1} \to \mathbb{R}$ be the consumer surplus, i.e. the expected maximum utility given by

$$\overline{CS}(\boldsymbol{\delta}) = \mathbb{E}\left(\max_{j \in \mathcal{J}} u_j\right).$$

By the Williams-Daly-Zachary theorem (McFadden, 1981), the conditional choice probabilities are equal to the derivatives of the consumer surplus, i.e. $P_j(\boldsymbol{\delta}) = \partial \overline{CS}(\boldsymbol{\delta}) / \partial \delta_j$. Define a function $\overline{\mathbf{H}} = \left(\overline{H}_0, \ldots, \overline{H}_J\right)$, with $\overline{H}_j : (0,\infty)^{J+1} \to (0,\infty)$ as the derivative of the exponentiated surplus with respect to its jth component, i.e.,

$$\overline{H}_j\left(e^{\delta}\right) = \frac{\partial e^{\overline{CS}(\boldsymbol{\delta})}}{\partial \delta_j} = P_j(\boldsymbol{\delta})\, e^{\overline{CS}(\boldsymbol{\delta})}, \quad j \in \mathcal{J}.$$

Summing over $k \in \mathcal{J}$ and using that probabilities sum to one, we can write the ARUM choice probabilities and the consumer surplus in terms of $\overline{\mathbf{H}}$ as

$$P_j(\boldsymbol{\delta}) = \frac{\overline{H}_j\left(e^{\delta}\right)}{\sum_{k \in \mathcal{J}} \overline{H}_k\left(e^{\delta}\right)}, \quad j \in \mathcal{J}, \tag{18}$$

[18] Note that income does not enter utility (17), which means that there is no income effect. This is equivalent to the case in which income enters linearly. The deterministic utilities, δ_j, are common across all consumers, which rules out heterogeneity in preferences apart from the random utilities ϵ_j.

and

$$\overline{CS}(\delta) = \ln\left(\sum_{k\in\mathcal{J}} \overline{H}_k\left(e^{\delta}\right)\right).$$

Lemma 6 in Appendix B shows that $\overline{\mathbf{H}}$ is invertible, with inverse $\overline{\mathbf{G}} = \overline{\mathbf{H}}^{-1}$, and that $\ln\overline{\mathbf{G}}$ is an inverse GIL demand. Then we can invert Equations (18) to obtain the inverse ARUM demands, which coincide with the inverse GIL demands (10) when $\mathbf{G} = \overline{\mathbf{G}}$,

$$\ln\overline{G}_j(\mathbf{s}) + c = \delta_j, \quad j \in \mathcal{J},$$

with $c = \overline{CS}(\delta)$.

Products are always substitutes in an ARUM. In contrast, some GIL models allow for complementarity and cannot therefore be rationalized by any ARUM. This is in particular the case of the IPDL model introduced in Section 3 and used in the empirical illustration in Section 4. We summarize the results as follows.

Proposition 3. The ARUM choice probabilities in Equation (18) coincide with the GIL demands defined by Equation (11) when $\mathbf{G} = \overline{\mathbf{G}} = \mathbf{H}^{-1} = \overline{\mathbf{H}}^{-1}$.

Then any ARUM is consistent with some GIL model. The converse does not hold, since some GIL models are not consistent with any ARUM.

Lastly, any GIL model is consistent with some perturbed utility model (PUM).[19] In a PUM, the consumer chooses a vector of choice probabilities $\mathbf{s} \in \operatorname{int}(\Delta)$ to maximize her utility function defined as the sum of an expected utility component and a concave and deterministic function of $\mathbf{s}$, labeled as perturbation. Specifically, the GIL model (16) can be rationalized by a PUM with utility given by

$$\sum_{j\in\mathcal{J}} \delta_j s_j - \sum_{j\in\mathcal{J}} s_j \ln G_j(\mathbf{s}),$$

without the explicit structure of income and prices. However, the converse does not hold. For example, for the concave perturbation function $\sum_{j\in\mathcal{J}} \ln(s_j)$, the corresponding candidate inverse GIL demand is $\ln G_j(\mathbf{s}) = \frac{1}{s_j}\ln\left(s_j\right)$, which is not homogeneous of degree one and thus is not an inverse GIL demand.

[19] See Hofbauer and Sandholm (2002), McFadden and Fosgerau (2012) and Fudenberg et al. (2015) for more details on PUM. PUM have been used to model optimization with effort (Mattsson and Weibull, 2002), stochastic choices (Swait and Marley, 2013; Fudenberg et al., 2015), and rational inattention (Matejka and McKay, 2015; Fosgerau et al., 2018). Allen and Rehbeck (2019) show that some PUM allow for complementarity

Proposition 3 shows that the choice probabilities generated by any ARUM can be derived from some GIL model. As the class of GIL models is a strict subset of the class of PUM models, we have therefore strengthened Hofbauer and Sandholm (2002)'s result that the choice probabilities generated by any ARUM can be derived from some PUM by showing that the GIL structure is sufficient to recover any ARUM.

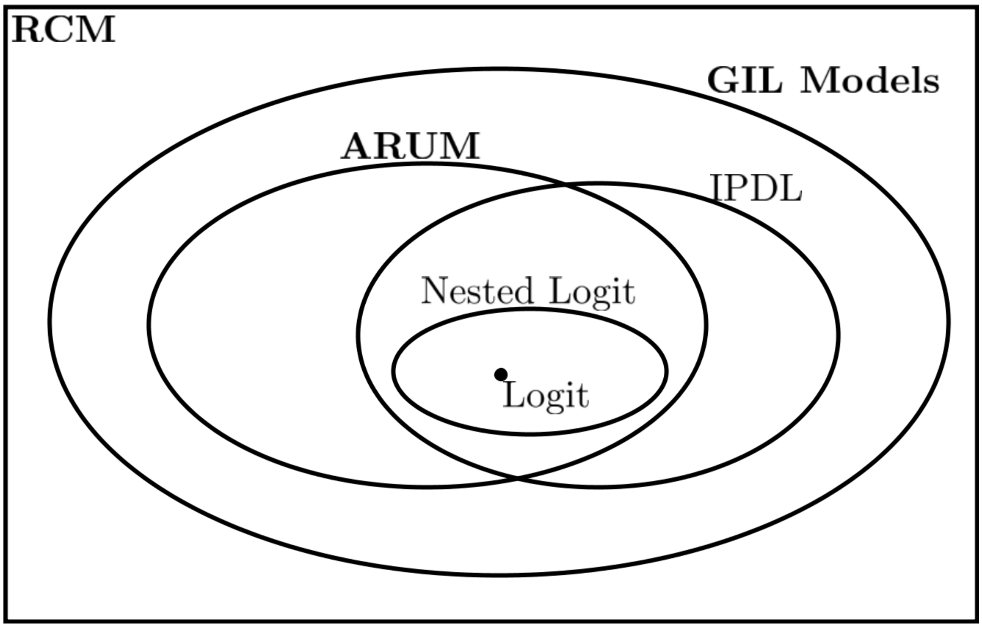
6.3 Overview of Relationships

The relationships between the GIL, IDPL, ARUM and RCM classes of models are illustrated in Figure 3.

We have established that any GIL model is an RCM. An example suffices to show that there are RCM that are not consistent with any GIL model. In particular, when $\ln G_j(\mathbf{s}) = \frac{1}{s_j}\ln\left(s_j\right)$, the direct utility (14) is consistent with a RCM but not with a GIL model.

As mentioned above, the IPDL model is a GIL model and admits the logit and nested logit models as special cases. We have also shown that any ARUM is observationally equivalent to some GIL model. However, the special case of IPDL model shows that the converse does not hold, since it allows for complementarity which is ruled out by any ARUM.

Figure 3: Relationships between RCM, ARUM and GIL models



Altogether, as Figure 3 shows, the class of GIL model is strictly larger than the class of ARUM, but strictly smaller than the class of RCM.

7 Conclusion

This paper has introduced the IPDL model, which is an inverse demand model for differentiated products that are segmented according to multiple dimensions. The IDPL model allows for more complex patterns of substitution than the nested logit model, it even allows for complementarity, while being easily estimated by linear regression using Berry (1994)'s method. The IDPL model provides an attractive modelling framework in applications where the priority is to maintain the computational simplicity of logit and nested logit models, while allowing more realistic patterns of substitution that do not constrain products to be substitutes.

The IDPL model belongs to the wider class of GIL models, which is a class of representative consumer models, large enough to comprise equivalents of all ARUM as well as models in which products may be complements. Finding that GIL demands are invertible even in the presence of complementarity extends the previous literature on invertibility of demand.

There is ample room for future research on the IDPL model and the more general GIL class of models. Generally, it is of interest to develop GIL models for various applications, exploiting the possibilities for constructing models with structures that are tailored to specific circumstances. On the methodological level, it is of interest to develop methods for estimating GIL models with individual-level data. Another issue is to determine conditions on the inverse GIL demand under which products are substitutes. Finally, the link to rational inattention, pointed out in Fosgerau et al. (2018), remains to be explored theoretically and empirically.

Appendices

A Proofs and Additional Results

A.1 Mathematical Notation

We use italics for scalar variables and real-valued functions, boldface for vectors, matrices and vector-valued functions, and calligraphic for sets. By default, vectors are column vectors: $\mathbf{s} = \left(s_0, \ldots, s_J\right)^\intercal \in \mathbb{R}^{J+1}$.

$\Delta \subset \mathbb{R}^{J+1}$ is the J-dimensional unit simplex: $\Delta = \left\{\mathbf{s} \in [0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$, and $\operatorname{int}(\Delta) = \left\{\mathbf{s} \in (0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$ is its interior.

Let $CS : \mathbb{R}^{J+1} \to \mathbb{R}$ be a function. Then, $\nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta})$, with elements j given by $\frac{\partial CS(\boldsymbol{\delta})}{\partial \delta_j}$, denotes its gradient with respect to the vector $\boldsymbol{\delta}$.

Let $\mathbf{G} = \left(G_0, \ldots, G_J\right) : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a vector function composed of functions $G_j : \mathbb{R}^{J+1} \to \mathbb{R}$. Its Jacobian matrix $\mathbf{J_G}(\mathbf{s})$ at $\mathbf{s}$ has elements ij given by $\frac{\partial G_i(\mathbf{s})}{\partial s_j}$.

A univariate function $\mathbb{R} \to \mathbb{R}$ applied to a vector is a coordinate-wise application of the function, e.g., $\ln(\mathbf{s}) = \left(\ln(s_0), \ldots, \ln\left(s_J\right)\right)$. $\mathbf{1}_J = (1, \ldots, 1)^\intercal \in \mathbb{R}^J$ is a vector consisting of ones and $\mathbf{I}_J \in \mathbb{R}^{J\times J}$ denotes the $J \times J$ identity matrix.

A.2 Preliminary Results

This section states some preliminary mathematical results that are used in the proofs below.

Lemma 1 (Euler equation for homogeneous functions)**.** Suppose that $\phi : (0,\infty)^{J+1} \to \mathbb{R}$ is homogeneous of degree one. Then

$$\phi(\mathbf{s}) = \sum_{i=0}^{J} \frac{\partial \phi(\mathbf{s})}{\partial s_i} s_i, \quad \text{for all } \mathbf{s} \in (0,\infty)^{J+1}.$$

Definition 2. A matrix $\mathbf{A} \in \mathbb{R}^{(J+1)\times(J+1)}$ is positive quasi-definite if its symmetric part, defined by $\frac{1}{2}(\mathbf{A}+\mathbf{A}^\top)$, is positive definite.

It follows that a symmetric and positive definite matrix is positive quasi-definite.

Lemma 2 (Gale and Nikaido 1965, Theorem 6)**.** If a differentiable mapping $\mathbf{F} : \Theta \to \mathbb{R}^{J+1}$, where Θ is a convex region (either closed or non-closed) of $\mathbb{R}^{J+1}$, has a Jacobian matrix that is everywhere quasi-definite in Θ, then $\mathbf{F}$ is injective on Θ.

Lemma 3 (Simon and Blume, 1994, Theorem 14.4)**.** Let $\mathbf{F} : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ and $\mathbf{G} : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be continuously differentiable functions. Let $\mathbf{y} \in \mathbb{R}^{J+1}$ and $\mathbf{x} = \mathbf{G}(\mathbf{y}) \in \mathbb{R}^{J+1}$. Consider the composite function

$$\mathbf{C} = \mathbf{F} \circ \mathbf{G} : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}.$$

The Jacobian matrix $\mathbf{J}_\mathbf{C}(\mathbf{y})$ is given by

$$\mathbf{J}_\mathbf{C}(\mathbf{y}) = \mathbf{J}_{\mathbf{F}\circ\mathbf{G}}(\mathbf{y}) = \mathbf{J}_\mathbf{F}(\mathbf{x})\mathbf{J}_\mathbf{G}(\mathbf{y}).$$

A.3 Properties of the IPDL Model

Let $\boldsymbol{\Theta}_d$ be the matrix encoding the grouping structure for dimension d with elements ij given by

$$(\boldsymbol{\Theta}_d)_{ij} = \begin{cases} 1, & \text{if } i \in \mathcal{G}_d(j), \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{G}_d(j)$ is the set of products that are grouped with product j in dimension d. Let $s_{\mathcal{G}_d(j)} = \sum_{k\in\mathcal{J}} (\boldsymbol{\Theta}_d)_{jk} s_k$ denote the market share of the group $\mathcal{G}_d(j)$.

Proposition 4. The IPDL model has the following properties.

1. The IIA property holds for products of the same type; but does not hold in general for products of different types.

2. The matrix of own- and cross-price derivatives is given by

$$\mathbf{J}_\sigma(\delta) = -\alpha\left(\boldsymbol{\Psi}\operatorname{diag}(\mathbf{s}) - \mathbf{s}\mathbf{s}^\top\right), \tag{19}$$

where

$$\boldsymbol{\Psi} = \left[\left(1 - \sum_{d=1}^{D} \mu_d\right)\mathbf{I}_{J+1} + \sum_{d=1}^{D} \mu_d \boldsymbol{\Theta}_d \mathbf{S}_{\mathcal{G}_d}\right]^{-1}, \tag{20}$$

where $\mathbf{S}_{\mathcal{G}_d}$ is the diagonal matrix with elements jj given by $\frac{s_j}{s_{\mathcal{G}_d(j)}}$ with $s_j = \sigma_j(\boldsymbol{\delta})$.

3. Products can be substitutes or complements.

Proof of Proposition 4

1. Using the relation (6) between indexes $\boldsymbol{\delta}$ and market shares $\mathbf{s}$, we obtain for any pair of products j and k that

$$\frac{\sigma_j(\boldsymbol{\delta})}{\sigma_k(\boldsymbol{\delta})} = \exp\left(\frac{\delta_j - \delta_k}{1 - \sum_{d=1}^{D} \mu_d} + \sum_{d=1}^{D} \frac{\mu_d}{1 - \sum_{d=1}^{D} \mu_d} \ln\left(\frac{\sigma_{\mathcal{G}_d(k)}(\boldsymbol{\delta})}{\sigma_{\mathcal{G}_d(j)}(\boldsymbol{\delta})}\right)\right). \tag{21}$$

For products j and k of the same type (i.e., with $\mathcal{G}_d(k) = \mathcal{G}_d(j)$ for all d), Equation (21) reduces to $\frac{\sigma_j(\boldsymbol{\delta})}{\sigma_k(\boldsymbol{\delta})} = \exp\left(\frac{\delta_j - \delta_k}{1 - \sum_{d=1}^{D} \mu_d}\right)$, which is independent of the characteristics or existence of all other products, i.e., IIA holds for products of the same type. When products are of different types, the ratio can depend on the characteristics of other products, which means that IIA does not hold in general.

2. Use Equation (25) in Proposition 5 below to show that the matrix of own- and cross-price derivatives is given by Equations (19) and (20).

3. Suppose there are 3 inside products and one outside good. Inside products are grouped according two dimensions. For the first dimension, product 1 is in one group, and products 2 and 3 are in a second group. For the second dimension, products 1 and 2 are in one group, and product 3 is in a second group.

   Using Equation (19), we show that

$$\frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_3} = -\alpha\left((1 - \mu_1 - \mu_2)(s_1 + s_2)(s_2 + s_3) - \mu_1 \mu_2 s_0 s_2\right),$$

   meaning that if

$$(1 - \mu_1 - \mu_2)(s_1 + s_2)(s_2 + s_3) - \mu_1 \mu_2 s_0 s_2 > 0,$$

   then products 1 and 3 are complements. □

A.4 Results for Section 5

Proof of Proposition 1 The function $\ln \mathbf{G}$ is differentiable on the convex region $\operatorname{int}(\Delta)$ of $\mathbb{R}^{J+1}$. In addition, $\mathbf{J}_{\ln \mathbf{G}}$ is positive quasi-definite on $\operatorname{int}(\Delta)$, since by assumption it is symmetric and positive definite on $\operatorname{int}(\Delta)$. Then $\ln \mathbf{G}$ is injective by Lemma 2. $\square$

Proposition 5. The GIL models defined by Equation (10) satisfy the following properties.

1. The market-specific constant c is equal to

$$c = \ln\left(\sum_{k\in\mathcal{J}} H_k\left(e^{\delta}\right)\right), \tag{22}$$

   where $\mathbf{H}(e^{\delta}) = (H_0(e^{\delta}), \dots, H_J(e^{\delta})) = \mathbf{G}^{-1}(e^{\delta})$.

2. The market shares functions are given by

$$\sigma_j(\delta) = \frac{H_j\left(e^{\delta}\right)}{\sum_{k\in\mathcal{J}} H_k\left(e^{\delta}\right)}, \quad j \in \mathcal{J}. \tag{23}$$

3. The Euler-type equation

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} = 1, \quad k \in \mathcal{J}, \quad \mathbf{s} \in \operatorname{int}(\Delta)$$

   holds and can be written in matrix form as

$$\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})\mathbf{s} = \mathbf{1}_{J+1}, \quad \mathbf{s} \in \operatorname{int}(\Delta). \tag{24}$$

4. Roy's identity implies that the consumer surplus is given by the convex function

$$CS(\delta) = \ln\left(\sum_{k\in\mathcal{J}} H_k\left(e^{\delta}\right)\right).$$

5. With $\mathbf{s} = \boldsymbol{\sigma}(\delta)$, the matrix of demand derivatives is given by

$$\mathbf{J}_{\sigma}(\delta) = [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{-1} - \mathbf{s}\mathbf{s}^{\intercal}, \tag{25}$$

which is symmetric and positive semi-definite, implying that GIL demands have symmetric cross effects and are non-decreasing in their own index.

Proof of Proposition 5

1. Exponentiating and applying $\mathbf{H}$ on both sides of Equation (10) leads to

$$\mathbf{s} = \mathbf{H}(e^{\delta}e^{-c}) = \mathbf{H}(e^{\delta})e^{-c}, \tag{26}$$

where the last equality uses the homogeneity of $\mathbf{H}$. Using that demands sum to 1 leads to Equation (22).

2. Combine Equations (22) and (26) and use $\sigma_j(\boldsymbol{\delta}) = s_j$ to obtain Equation (23).

3. Note that

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} = \sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_k(\mathbf{s})}{\partial s_j} = \frac{\sum_{j\in\mathcal{J}} s_j \frac{\partial G_k(\mathbf{s})}{\partial s_j}}{G_k(\mathbf{s})} = \frac{G_k(\mathbf{s})}{G_k(\mathbf{s})} = 1,$$

where the first equality relies on the symmetry of the Jacobian of $\ln \mathbf{G}$ and the third equality uses the Euler equation for the homogeneous function $\mathbf{G}$.

4. We verify that Roy's identity holds. Set $\boldsymbol{\delta} = \ln \mathbf{G}(\mathbf{s})$. Then $(\ln \mathbf{G})^{-1}(\boldsymbol{\delta}) = \mathbf{H} \circ \exp(\boldsymbol{\delta}) = \mathbf{s}$, and by Lemma 3,

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

By assumption, the Jacobian $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is positive definite and symmetric. Then its inverse $\mathbf{J}_{\mathbf{H}\circ\exp}(\boldsymbol{\delta})$ exists and is symmetric, i.e.,

$$\frac{\partial H_i\left(e^{\delta}\right)}{\partial \delta_j} = \frac{\partial H_j\left(e^{\delta}\right)}{\partial \delta_i}.$$

Then Roy's identity can be verified via

$$\begin{aligned}\frac{\partial CS\left(e^{\delta}\right)}{\partial \delta_i} &= \frac{\sum_{k\in\mathcal{J}} \frac{\partial H_k(e^{\delta})}{\partial \delta_i}}{\sum_{j\in\mathcal{J}} H_j(e^{\delta})} = \frac{\sum_{k\in\mathcal{J}} \frac{\partial H_i(e^{\delta})}{\partial \delta_k}}{\sum_{j\in\mathcal{J}} H_j(e^{\delta})},\\ &= \frac{\sum_{k\in\mathcal{J}} \frac{\partial H_i(e^{\delta})}{\partial e^{\delta_k}} e^{\delta_k}}{\sum_{j\in\mathcal{J}} H_j(e^{\delta})} = \frac{H_i\left(e^{\delta}\right)}{\sum_{j\in\mathcal{J}} H_j(e^{\delta})} = \sigma_i(\boldsymbol{\delta}),\end{aligned}$$

where the second equality uses symmetry of $\mathbf{J}_{\mathbf{H}\circ\exp}(\boldsymbol{\delta})$ and the fourth equality uses the Euler equation for the homogeneous function $\mathbf{H}$.

Convexity of the consumer surplus follows by property 5 since the Hessian, $\mathbf{J}_{\sigma}(\boldsymbol{\delta})$, is positive semidefinite.

5. Differentiate $\delta_j = \ln G_j(\mathbf{s}) + CS(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ to find that

$$\mathbf{I}_{J+1} = \mathbf{J}_{\ln \mathbf{G}}(\mathbf{s}) \mathbf{J}_{\sigma}(\boldsymbol{\delta}) + \mathbf{1}_{J+1}\mathbf{s}^{\intercal},$$

where $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$. Solving for $\mathbf{J}_{\sigma}(\boldsymbol{\delta})$, we obtain that

$$\mathbf{J}_{\sigma}(\boldsymbol{\delta}) = [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{-1}\left[\mathbf{I} - \mathbf{1}_{J+1}\mathbf{s}^{\intercal}\right] = [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{-1} - [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{-1}\mathbf{1}_{J+1}\mathbf{s}^{\intercal},$$

since $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is invertible.

Finally, use Equation (24) to show that $[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{-1}\mathbf{1}_{J+1}\mathbf{s}^{\intercal} = \mathbf{s}\mathbf{s}^{\intercal}$. Then $\mathbf{J}_{\sigma}(\boldsymbol{\delta})$ is symmetric.

As $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is positive definite, the square-root matrix $[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{1/2}$ exists and is also positive definite. Then

$$[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{1/2}\mathbf{J}_{\sigma}(\boldsymbol{\delta})[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{1/2} = [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{-1/2}(\mathbf{I}_{J+1} - \mathbf{1}_{J+1}\mathbf{s}^{\intercal})[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})]^{1/2},$$

is symmetric and idempotent and hence positive semidefinite. Then also $\mathbf{J}_{\sigma}(\boldsymbol{\delta})$ is positive semidefinite. □

A.5 Results for Section 6

Representative Consumer Model

Lemma 4. Let $\ln \mathbf{G}$ be an inverse GIL demand. Then the function $\mathbf{s} \to -\mathbf{s}^{\intercal}\ln \mathbf{G}(\mathbf{s}) = -\sum_{j\in\mathcal{J}} s_j \ln G_j(\mathbf{s})$ is strictly concave on $\operatorname{int}(\Delta)$.

Proof of Lemma 4 Consider $\mathbf{s} \in \operatorname{int}(\Delta)$. By property 3 of Proposition 5, the Hessian of $-\mathbf{s}^{\intercal}\ln \mathbf{G}(\mathbf{s})$ is $-\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$, which is negative definite by assumption. □

Proof of Proposition 2 Consider the representative consumer maximizing utility (14) subject to constraints (15). The budget constraint is always binding since $\alpha > 0$ and $y > \max_{j\in\mathcal{J}} p_j$. Substituting the budget constraint into the direct utility

(14), the representative consumer then chooses $\mathbf{s} \in \Delta$ to maximize

$$u(\mathbf{s}) = \alpha y + \sum_{j \in \mathcal{J}} \delta_j s_j - \sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s}),$$

where $\delta_j = v_j - \alpha p_j$.

The Lagrangian of the utility maximization program given by

$$\mathcal{L}(\mathbf{s}, \lambda) = u(\mathbf{s}) + \lambda \left(1 - \sum_{j \in \mathcal{J}} s_j\right),$$

yields $\sum_{j \in \mathcal{J}} s_j = 1$ and the first-order conditions

$$\delta_j - \ln G_j(\mathbf{s}) - \sum_{k \in \mathcal{J}} s_k \frac{\partial \ln G_k(\mathbf{s})}{\partial s_j} - \lambda = 0, \quad j \in \mathcal{J}.$$

By property 3 of Proposition 5, the first-order conditions can be simplified to

$$\delta_j - \ln G_j(\mathbf{s}) - 1 - \lambda = 0, \quad j \in \mathcal{J}.$$

The first-order condition for an interior solution has a unique solution, since the objective is strictly concave by Lemma 4, hence the utility maximizing demands exist uniquely.

Setting $c = 1 + \lambda$, one obtains

$$\ln G_j(\mathbf{s}) + c = \delta_j,$$

which then shows that the representative consumer model leads to the inverse GIL demand. □

Additive Random Utility Model

Since shifting all the δ_j's by a constant amount $c \in \mathbb{R}$ shifts the maximum expected utility $\overline{CS}$ by the same amount and does not affect choice probabilities $\mathbf{P}$, we may use the normalization $\sum_{j \in \mathcal{J}} \delta_j = 0$, i.e., we consider at no loss of generality the restrictions of $\overline{G}$ and $\mathbf{P}$ to $\Lambda = \left\{\boldsymbol{\delta} \in \mathbb{R}^{J+1} : \sum_{j \in \mathcal{J}} \delta_j = 0\right\}$. The following lemma collects some properties of the expected maximum utility $\overline{CS}$.

Lemma 5. The expected maximum utility $\overline{CS}$ has the following properties.

1. It is twice continuously differentiable, convex and finite everywhere.
2. It satisfies the homogeneity property
$$\overline{CS}\left(\boldsymbol{\delta}+c\mathbf{1}_{J+1}\right)=\overline{CS}\left(\boldsymbol{\delta}\right)+c,\quad c\in\mathbb{R}. \tag{27}$$
3. Its Hessian is positive definite on Λ.
4. It is given in terms of the expected residual of the maximum utility product by
$$\overline{CS}\left(\boldsymbol{\delta}\right)=\sum_{j\in\mathcal{J}}P_j\left(\boldsymbol{\delta}\right)\delta_j+\mathbb{E}\left(\varepsilon_{j^*}|\boldsymbol{\delta}\right), \tag{28}$$
where j^* is the index of the chosen product.

Proof of Lemma 5 McFadden (1981) establishes convexity, finiteness, and the homogeneity property (27). He also shows the existence of all mixed partial derivatives up to order $J\geq 2$, meaning that all second order mixed partial derivatives are continuous. Hofbauer and Sandholm (2002) show that the Hessian of $\overline{CS}$ is positive definite on Λ.

Let j^* be the index of the chosen product. Property (28) follows from
$$\begin{aligned}\overline{CS}\left(\boldsymbol{\delta}\right)&=\sum_{j\in\mathcal{J}}\mathbb{E}\left(\max_{j\in\mathcal{J}}\left\{\delta_j+\varepsilon_j\right\}|j^*=j,\boldsymbol{\delta}\right)P_j\left(\boldsymbol{\delta}\right),\\&=\sum_{j\in\mathcal{J}}\left(\delta_j+\mathbb{E}\left(\varepsilon_{j^*}|j^*=j,\boldsymbol{\delta}\right)P_j\left(\boldsymbol{\delta}\right)\right),\\&=\sum_{j\in\mathcal{J}}P_j\left(\boldsymbol{\delta}\right)\delta_j+\mathbb{E}\left(\varepsilon_{j^*}|\boldsymbol{\delta}\right),\end{aligned}$$
where the first equality uses the law of iterated expectations. □

It is well-known in the convex analysis literature that, for the logit model, the convex conjugate of the negative Shannon entropy $-\overline{CS}^*(\mathbf{s})=\sum_{j\in\mathcal{J}}s_j\ln\left(s_j\right)$ is the log-sum $\overline{CS}\left(\boldsymbol{\delta}\right)=\ln\left(\sum_{j\in\mathcal{J}}e^{\delta_j}\right)$ (see e.g., Boyd and Vandenberghe, 2004). Fosgerau et al. (2018) extend this result to a class of "generalized entropies" which has the Shannon entropy as special case. See also Matejka and McKay (2015), Chiong et al. (2016) and Galichon and Salanié (2015) who use convex analysis in different contexts.

Lemma 6. The function $\overline{\mathbf{H}}$ is invertible, and its inverse $\overline{\mathbf{G}} = \overline{\mathbf{H}}^{-1}$ is an inverse GIL demand.

Lemma 6 is proved in Fosgerau et al. (2018) in a very similar setting. The proof provided here applies to the exact setting of the current paper and has independent value by being simpler.

Proof of Lemma 6 We first show that $\overline{\mathbf{H}}$ is injective. Note that $\overline{\mathbf{H}}$ is differentiable. Consider the function $\delta \to \overline{\mathbf{H}}\left(e^{\delta}\right)$. Its Jacobian is positive definite on Λ since it has elements ij given by

$$\left\{e^{\overline{CS}(\delta)} \frac{\partial \overline{CS}(\delta)}{\partial \delta_i} \frac{\partial \overline{CS}(\delta)}{\partial \delta_j}\right\} + \left\{e^{\overline{CS}(\delta)} \frac{\partial^2 \overline{CS}(\delta)}{\partial \delta_i \partial \delta_j}\right\},$$

where the first term is a positive semi-definite matrix and where, by property 3 of Lemma 5, the second term is a positive definite matrix on Λ. As it is also symmetric, it follows that the Jacobian is positive quasi-definite. Then $\overline{\mathbf{H}}$ is invertible by Lemma 2. By Norets and Takahashi (2013), the range of $\overline{\mathbf{H}}$ is $\mathrm{int}(\Delta)$, which then is the domain of the inverse function $\overline{\mathbf{H}}^{-1}$.

We now show that $\ln \overline{\mathbf{G}}$ is an inverse GIL demand. Note that $\overline{\mathbf{G}}$ is linearly homogeneous and that, as shown above, the Jacobian of $\overline{\mathbf{H}}$ is symmetric and positive definite. Then, by Lemma 3, the same is true for the Jacobian of $\ln \overline{\mathbf{G}}$. □

B Data

Databases We use data from the Dominick's Database made available by the James M. Kilts Center, University of Chicago Booth School of Business. This is weekly store-level scanner data, comprising information on 30 categories of packaged products at the UPC level for all Dominick's Finer Foods chain stores in the Chicago metropolitan area over the period 1989-1997. For the application, we consider the RTE cereals category during the period 1991–1992.

We supplement the data with the nutrient content of the RTE cereals using the USDA Nutrient Database for Standard Reference. This dataset is made available by the United States Department of Agriculture and provides the nutrient content of more than 8,500 different foods including RTE cereals (in particular, we use releases SR11 (year 1996) and SR16 (year 2004) for sugar). We have collected six characteristics: fiber, sugar, lipid and protein in g/serve, energy in kcal/serve, and sodium in mg/serve. We also supplement the data with monthly sugar prices from the website www.indexmundi.com to form cost-based instruments.

Markets, Products, Market shares and Prices We aggregate UPCs into products (e.g., Kellogg's Special K), so that different size boxes are considered one product, where a product is a cereal (e.g., Special K) associated to its brand (e.g., Kellogg's). We focus attention on the top 50 products in terms of sales, which account for 73 percent of sales of the category in the sample we use.

We define a market as a store-month pair. Following Nevo (2001), we define market shares of the inside products by converting volume sales into number of servings sold, and then by dividing it by the total potential number of servings at a store in a given month.

To compute the total potential number of servings at a store in a given month, we assume that (i) an individual in a household consumes around 15 servings per month, and (ii) consumers visit stores twice a week. Indeed, according to USDA's Economic Research Service, per capita consumption of RTE cereals was equal to around 14 pounds (that is, about 6350 grammes) in 1992, which is equivalent to 15 servings per month (without loss of generality, we assume that a serving weight is equal to 35 grammes). Then, the potential (month-store) market size (in servings) is computed as the weekly average number of households which visited that store in that given month, times the average household size for that store, times the number of servings an individual consumes in a month. The market share of the outside good is then the difference between one and the sum of the inside products market shares. As a robustness check, we have also estimated the models with the alternative assumption that consumers visit stores once a week; results do not change significantly.

Lastly, following Nevo (2001), we compute the price of a serving by dividing the dollar sales by the number of servings sold, where the dollar sales reflect the price consumers paid; we also convert the six nutrients into nutrients for a serving.

Supplements

Notation We use italics for scalar variables and real-valued functions, boldface for vectors, matrices and vector-valued functions, and calligraphic for sets. By default, vectors are column vectors: $\mathbf{s} = \left(s_0, \ldots, s_J\right)^\intercal \in \mathbb{R}^{J+1}$.

$\Delta_J \subset \mathbb{R}^{J+1}$ is the J-dimensional unit simplex: $\Delta_J = \left\{\mathbf{s} \in [0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$, and $\operatorname{int}\left(\Delta_J\right) = \left\{\mathbf{s} \in (0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$ is its interior, where $\mathcal{J} = \{0, 1, \ldots, J\}$.

Let $CS : \mathbb{R}^{J+1} \to \mathbb{R}$ be a function. Then, $\nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta})$, with elements j given by $\frac{\partial CS(\delta)}{\partial \delta_j}$, denotes its gradient with respect to the vector $\boldsymbol{\delta}$.

Let $\mathbf{G} = \left(G_0, \ldots, G_J\right) : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a vector function composed of functions $G_j : \mathbb{R}^{J+1} \to \mathbb{R}$. Its Jacobian matrix $\mathbf{J_G}(\mathbf{s})$ at $\mathbf{s}$ has elements ij given by $\frac{\partial G_i(\mathbf{s})}{\partial s_j}$.

A univariate function $\mathbb{R} \to \mathbb{R}$ applied to a vector is a coordinate-wise application of the function, e.g., $\ln(\mathbf{s}) = \left(\ln(s_0), \ldots, \ln\left(s_J\right)\right)$. $\mathbf{1}_J = (1, \ldots, 1)^\intercal \in \mathbb{R}^J$ is a vector consisting of ones and $\mathbf{I}_J \in \mathbb{R}^{J\times J}$ denotes the $J \times J$ identity matrix. Let $|\tilde{\mathbf{s}}| = \sum_{j\in\mathcal{J}} |\tilde{s}_j|$ denotes the 1-norm of vector $\tilde{\mathbf{s}}$.

1 Simulation Results for the IPDL Model

Let $\boldsymbol{\Theta}_d$ be the matrix encoding the grouping structure for dimension d, with elements ij given by

$$(\boldsymbol{\Theta}_d)_{ij} = \begin{cases} 1, & \text{if } i \in \mathcal{G}_d(j), \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{G}_d(j)$ is the set of products that are grouped with product j in dimension d. Let $s_{\mathcal{G}_d(j)} = \sum_{k\in\mathcal{J}} (\boldsymbol{\Theta}_d)_{jk} s_k$ be the market share of $\mathcal{G}_d(j)$.

Recall that the matrix of own- and cross-price derivatives for the IPDL model is

$$\mathbf{J}_\sigma(\delta) = -\alpha\left(\boldsymbol{\Psi}\operatorname{diag}(\mathbf{s}) - \mathbf{s}\mathbf{s}^\intercal\right),$$

where

$$\boldsymbol{\Psi} = \left[\left(1 - \sum_{d=1}^{D} \mu_d\right)\mathbf{I}_{J+1} + \sum_{d=1}^{D} \mu_d \boldsymbol{\Theta}_d \mathbf{S}_{\mathcal{G}_d}\right]^{-1},$$

and where $\mathbf{S}_{\mathcal{G}_d}$ is the diagonal matrix with elements jj given by $\frac{s_j}{s_{\mathcal{G}_d(j)}}$ with $s_j = \sigma_j(\boldsymbol{\delta})$. We cannot obtain closed-form formulae for the entries of the matrix of own- and cross-price derivatives. We therefore perform simulations to better understand the substitution patterns of the IPDL model.

Simulated Data We simulate

- A market with 20 inside products and an outside good;
- 20 different grouping structures (i.e. allocations of products in groups) along 3 dimensions, and with 3 groups per dimension. We obtain a grouping structure by simulating a 20×3 matrix of random numbers following a generalized Bernoulli distribution;
- 20 different vectors of grouping parameters $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_3)$. We obtain a vector of $\boldsymbol{\mu}$ by simulating a 4-vector of uniformly distributed random numbers, where the first element is μ_0, then normalizing so that $\boldsymbol{\mu} \in \text{int}(\Delta_3)$;
- 20 different vectors of market shares $\mathbf{s} = (s_0, \ldots, s_{20})$. We obtain a vector of market shares by simulating a 21-vector of uniformly distributed random numbers, where the first element is s_0, then by normalizing the vector of market shares of inside products so that $\mathbf{s} \in \text{int}(\Delta_{20})$.

This way of normalizing ensures that we simulate markets with very low and very high values for μ_0 and s_0. Combining the grouping structures, the grouping parameters, and the market shares, we form $8,000$ markets. The following table gives summary statistics on the simulated data.

Table S1: Summary Statistics on the Simulated Data

Variable	Mean	Min	Max
s_0	0.5253	0.0064	0.9906
s_j	0.0158	3e-06	0.0697
μ_0	0.4662	0.0697	0.9532
μ_1	0.2014	0.0135	0.8480
μ_2	0.1420	0.0175	0.4036
μ_3	0.1904	0.0059	0.5212

Grouping Structure Table S2 shows the distribution of the own- and cross-price derivatives according to the number of common groups.

Own-price elasticities are always negative, while cross-price elasticities can be either negative (complementarity) or positive (substitutability). Products of the same type are always substitutes. Products that are very similar (i.e., that are grouped together according to all dimensions, but one) are also always substitutes. Products that are very different can be either substitutes or complements. Products are less likely to be substitutes as they become more different.

Table S2: Distribution of Price Derivatives by Number of Common Groups

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

Notes: Column "$\mathbf{J}_\sigma > 0$" gives the percentage of positive cross-price elasticities according to the number of common groups (e.g., the row "2" concerns products that are grouped together into 2 groups). Column "Freq." gives the frequencies (in percentage) of the cross-price elasticities (e.g., 4.85 percent of the cross-price elasticities involve products of the same type).

Grouping Parameters Table S3 shows the distribution of cross-price derivative according to the proximity of products into the characteristics space used to form product types, as measured by the sum of grouping parameters $\mu_{jk} = \sum_{d=1}^{3} \mu_d \mathbf{1}\{j \in \mathcal{G}_d(k)\}$ for two products j and k.

As the parameter μ_{jk} becomes larger, we observe that (i) the derivatives increase in values, and that (ii) the share of substitutes increases. This is because higher μ_d means that products of the same group in dimension d become more similar.

Table S3: Percentage of Substitutes according to the Value of μ_{jk}

μ_{jk}	$\mathbf{J}_\sigma > 0$	Median	Min	Max
$[0, 0.1[$	65.60%	0.0000	-0.1539	0.0286
$[0.1, 0.2[$	96.37%	0.0002	-0.0538	0.1462
$[0.2, 0.3[$	93.52%	0.0003	-0.1114	0.1670
$[0.3, 0.4[$	94.16%	0.0007	-0.0673	0.2082
$[0.4, 0.5[$	93.89%	0.0009	-0.0432	0.2049
$[0.5, 1[$	100.00%	0.0020	1e-08	0.3100

Summary In the IPDL model,

1. (Grouping structure) Products of the same type are always substitutes. Products of different types may be substitutes or complements, depending on the degree of closeness between products as measured by the value of the parameters μ_d and by the closeness of the products into the characteristics space used to form product types. The closer two products are, the more likely they are to be substitutes.

2. (Grouping parameters) The size of the cross-derivatives depends on the degree of closeness. The closer two products are, the higher is their cross-derivative.

2 Construction of GIL Models

In this section, we provide a range of general methods for building members of the class of GIL models, along with illustrative examples. They allow us to obtain alternative models to the logit and nested logit models that have interesting features: some of them can accommodate complementarity, others have closed form for both the demands and their inverse.

Definition 1. An inverse GIL demand is a function $\ln \mathbf{G}$, where $\mathbf{G} : (0,\infty)^{J+1} \to (0,\infty)^{J+1}$ is homogeneous of degree one and where the Jacobian $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is positive definite and symmetric.

Definition 2. An almost inverse GIL demand is a function that satisfies the requirements for being an inverse GIL demand, except the Jacobian $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is only required to be positive semi-definite rather than positive definite.

2.1 General Methods and Illustrative Examples

The first result in this section shows that averaging an almost inverse GIL demand with an inverse GIL demand yields a new inverse GIL demand.

Proposition S1 (Averaging)**.** Let $\mathbf{G}_k, k \in \{1,\ldots,K\}$, be almost inverse GIL demands with at least one being an inverse GIL demand. Let $(\alpha_1,\ldots,\alpha_K) \in \mathrm{int}(\Delta_{K-1})$. Then

$$\ln \mathbf{G}(\mathbf{s}) = \sum_{k=1}^{K} \alpha_k \ln \mathbf{G}_k(\mathbf{s})$$

is an inverse GIL demand.

Proof of Proposition S1 The function $\mathbf{G}$ is homogeneous of degree one since for $\lambda > 0$,

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

where the second equality uses the homogeneity of the functions G_k and the fourth equality uses the restrictions on parameters $\sum_{k=1}^{K} \alpha_k = 1$.

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

Proposition S1 leads to the following corollary.

Corollary S1 (General grouping)**.** Let $\mathcal{G} \subseteq 2^{\mathcal{J}}$ be a finite set of groups with associated parameters μ_g, where $\mu_{0j} + \sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g = 1$ for all $j \in \mathcal{J}$ with $\mu_g \geq 0$ for all $g \in \mathcal{G}$ and $\mu_{0j} > 0$ for all $j \in \mathcal{J}$. Let $\ln \mathbf{G} = (\ln G_0, \ldots, \ln G_J)$ be given by

$$\ln G_j(\mathbf{s}) = \mu_{0j} \ln\left(s_j\right) + \sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g \ln\left(\sum_{i \in g} s_i\right).$$

Then $\ln \mathbf{G}$ is an inverse GIL demand.

Proof of Corollary S1 Let $T_j^0(\mathbf{s}) = s_j$ and for each $g \in \mathcal{G}$, $\mathbf{T}^g = \left(T_1^g, \ldots, T_J^g\right)$ with $T_j^g(\mathbf{s}) = \left(\sum_{i \in g} s_i\right)^{\mathbf{1}\{j \in g\}}$. The Jacobian of $\ln \mathbf{T}^g$ has elements jk given by $\frac{\mathbf{1}\{j \in g\}\mathbf{1}\{k \in g\}}{\sum_{i \in g} s_i}$, and thus $\mathbf{J}_{\ln \mathbf{T}^g} = \frac{\mathbf{1}_g \mathbf{1}_g^\intercal}{\sum_{i \in g} s_i}$ where $\mathbf{1}_g = (\mathbf{1}\{1 \in g\}, \ldots, \mathbf{1}\{J \in g\})^\intercal$. Each $\mathbf{T}^g$ is an almost inverse GIL demand while $\mathbf{T}_0$ is the logit inverse demand. Lastly, $\sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g + \mu_{0j} = 1$. Then the conditions for application of Proposition S1 are fulfilled. □

The grouping structure in Corollary S1 is arbitrary and therefore allows the grouping structure that defines the IPDL model. The presence of the logit inverse demand, due to $\mu_0 > 0$, ensures that the Jacobian $\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s})$ is always positive definite and hence that the inverse demand is indeed invertible.

If the outside good 0 belongs only to one group and is the only member of that group, then $\ln G_0(\mathbf{s}) = \ln(s_0) = \delta_0 + c$. Setting $\delta_0 = 0$ and assuming a linear index, the model of Corollary S1 boils down to the linear regression model

$$\ln\left(\frac{s_j}{s_0}\right) = \mathbf{x}_j \beta - \alpha p_j + \sum_{g \in \mathcal{G}(j)} \mu_g \ln\left(\sum_{k \in g} s_k\right) + \xi_j, \quad j = 1, \ldots, J.$$

The following proposition shows how an inverse GIL demand can be transformed into another inverse GIL demand by application of a location shift and a matrix with non-negative elements that sum to one across rows and columns. Let unnormalized demands $\tilde{\mathbf{s}}$ be demands obtained before normalizing their sum to one, i.e., $\mathbf{s} = \tilde{\mathbf{s}} / |\tilde{\mathbf{s}}|$.

Proposition S2 (Transformation)**.** Let $\mathbf{T}$ be an inverse GIL demand and $\mathbf{m} \in \mathbb{R}^{J+1}$ be a location shift vector. Let $\mathbf{A} \in \mathbb{R}^{(J+1)\times(J+1)}$ be an invertible matrix such that

$a_{ij} \geq 0$ and $\sum_{i\in\mathcal{J}} a_{ij} = \sum_{j\in\mathcal{J}} a_{ij} = 1$. Then the function $\ln \mathbf{G}$ given by

$$\ln \mathbf{G}(\mathbf{s}) = \mathbf{A}^{\intercal}\left[\ln\left(\mathbf{T}\left(\mathbf{A}\mathbf{s}\right)\right)\right] + \mathbf{m} \tag{1}$$

is an inverse GIL demand, and the corresponding unnormalized demands are given by

$$\tilde{\mathbf{s}} = \mathbf{A}^{-1}\mathbf{T}^{-1}\left(\exp\left[\left(\mathbf{A}^{\intercal}\right)^{-1}\left(\boldsymbol{\delta} - \mathbf{m}\right)\right]\right). \tag{2}$$

Proof of Proposition S2 The function $\mathbf{G}$ defined by Equation (1) is homogeneous of degree one since for $\lambda > 0$,

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

where the second equality uses the homogeneity of $\mathbf{T}$, and the third equality uses the feature that columns of $\mathbf{A}$ sum to 1.

The Jacobian of $\ln \mathbf{G}$ is $\mathbf{J}_{\ln\mathbf{G}}(\mathbf{s}) = \mathbf{A}^{\intercal}\mathbf{J}_{\ln\mathbf{T}}(\mathbf{s})\mathbf{A}$, which is symmetric and positive definite. Unnormalized demands (2) follow from solving $\ln\mathbf{G}\left(\tilde{\mathbf{s}}\right) = \boldsymbol{\delta}$. □

Proposition S2 provides models where both demand and inverse demand have closed form, as it is the case of the logit and nested logit models. We illustrate this proposition with an inverse GIL demand that allows for complementarity. Let $J + 1 = 3$, $\mathbf{m} = \mathbf{0}$, $\mathbf{T}\left(\mathbf{s}\right) = \mathbf{s}$, and

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

with $p < 0.5$. Then we obtain that

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

so that

$$s_3 = \sigma_3\left(\boldsymbol{\delta}\right) = \frac{e^{\delta_3}}{e^{\frac{p}{2p-1}\delta_1 - \frac{1-p}{2p-1}\delta_2} + e^{\frac{p}{2p-1}\delta_2 - \frac{1-p}{2p-1}\delta_1} + e^{\delta_3}},$$

and $\frac{\partial \sigma_3(\delta)}{\partial \delta_1} > 0$ if and only if $\delta_2 - \delta_1 > (2p-1)\ln\left(\frac{1-p}{p}\right)$.

2.2 Zero Demands

The constructions above rule out zero demands (this is also the case of the models discussed in the main text). The following proposition shows how we can build models that allow zero demands by slightly modifying Proposition S1 and applying it to functions $\mathbf{G}$ defined on $[0,\infty)^{J+1}$ instead of just $(0,\infty)^{J+1}$.

Proposition S3 (Invertible grouping). Let $\mathcal{G} = \left\{g_0, \ldots, g_J\right\}$ be a finite set of $J+1$ groups (i.e., the number of groups is equal to the number of products). Let $\mu_g > 0$, for all $g \in \mathcal{G}$, be the associated parameters, where $\sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g = 1$ for all $j \in \mathcal{J}$. Let $\mathbf{G} = (G_0, \ldots, G_J) : [0,\infty)^{J+1} \to (0,\infty)^{J+1}$ be given by

$$\ln G_j(\mathbf{s}) = \sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g \ln\left(\sum_{i \in g} s_i\right). \tag{3}$$

Let $\mathbf{W} = \operatorname{diag}\left(\mu_{g_0}, \ldots, \mu_{g_J}\right)$ and let $\mathbf{M} \in \mathbb{R}^{(J+1)\times(J+1)}$ with entries $M_{jk} = \mathbf{1}_{\{j \in g_k\}}$ (where rows correspond to products and columns to groups). If $\mathbf{M}$ is invertible, then $\ln \mathbf{G}$ has all the properties of an inverse GIL demand, except that it is defined on Δ_J, and the unnormalized demands satisfy

$$\boldsymbol{\delta} = \ln \mathbf{G}(\tilde{\mathbf{s}}) \Leftrightarrow \tilde{\mathbf{s}} = (\mathbf{M}^\intercal)^{-1} \exp\left(\mathbf{W}^{-1}\mathbf{M}^{-1}\boldsymbol{\delta}\right).$$

Proof of Proposition S3 Following the proof of Proposition S1, the function $\mathbf{G}$ given by Equation (3) clearly has all the properties of an almost inverse GIL demand. Thus, it remains to show that the Jacobian of $\ln \mathbf{G}$ is positive definite if $\mathbf{M}$ is invertible. Observe that

$$\begin{aligned}
\ln G_j(\mathbf{s}) &= \sum_{k \in \mathcal{J}} \mu_{g_k} \mathbf{1}\{j \in g_k\} \ln\left(\sum_{i \in g_k} s_i\right) \\
&= \sum_{k \in \mathcal{J}} \mu_{g_k} \mathbf{1}\{j \in g_k\} \ln\left(\sum_{i \in \mathcal{J}} \mathbf{1}\{i \in g_k\} s_i\right),
\end{aligned}$$

and, in turn, that

$$\frac{\partial \ln G_j(\mathbf{s})}{\partial s_l} = \sum_{k \in \mathcal{J}} \mu_{g_k} \frac{\mathbf{1}\{j \in g_k\}\mathbf{1}\{l \in g_k\}}{\sum_{i \in g_k} s_i},$$

which can be expressed in matrix form as

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

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

Lastly, unnormalized demands solve $\ln \mathbf{G}(\tilde{\mathbf{s}}) = \mathbf{M}\mathbf{W}\ln(\mathbf{M}^{\intercal}\tilde{\mathbf{s}}) = \boldsymbol{\delta}$ and, with $\mathbf{M}$ invertible, are given by $\tilde{\mathbf{s}} = (\mathbf{M}^{\intercal})^{-1}\exp\left(\mathbf{W}^{-1}\mathbf{M}^{-1}\boldsymbol{\delta}\right)$. □

As it is illustrated in the following example and as it is the case in ARUM where error terms have bounded support, Proposition S3 allows for zero demands when there is no degenerate group (i.e, a group containing only one product). Note that this proposition also allows to build models with closed form for both the demands and their inverses.

Define groups from the symmetric matrix $\mathbf{M}$ with entries $M_{ij} = \mathbf{1}_{\{i \neq j\}}$, so that each product belongs to $J+1$ groups. Its inverse, $\mathbf{M}^{-1}$, has entries ij equal to $\frac{1}{J+1} - 1_{\{i=j\}}$.

Let $\mu_g = 1/(J+1)$ for each group $g = 0, \ldots, J$. Then the unnormalized demands are given by $\tilde{\mathbf{s}} = (\mathbf{M})^{-1}\exp\left[(J+1)\mathbf{M}^{-1}\boldsymbol{\delta}\right]$ and lead to the following demands

$$\sigma_i(\boldsymbol{\delta}) = \frac{\tilde{s}_i}{\sum_{j \in \mathcal{J}} \tilde{s}_j} = \frac{\sum_{j \in \mathcal{J}} e^{-(J+1)\delta_j} - (J+1)e^{-(J+1)\delta_i}}{\sum_{j \in \mathcal{J}} e^{-(J+1)\delta_j}}. \tag{4}$$

Demands (4) are non-negative only for values of $\boldsymbol{\delta}$ within some set. To ensure positive demands, it is sufficient to average with the simple inverse logit demand. Demands (4) are not consistent with any ARUM since they do not exhibit the feature of the ARUM that the mixed partial derivatives of $\sigma_i(\boldsymbol{\delta})$ alternate in sign. Indeed, products are substitutes

$$\frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial \delta_2} = -J^2 e^{-J(\delta_1+\delta_2)} \Big/ \left(\sum_{j \in \mathcal{J}} e^{-J\delta_j}\right)^2 < 0,$$

but

$$\frac{\partial^2 \sigma_1(\boldsymbol{\delta})}{\partial \delta_2 \partial \delta_3} = -2J^3 e^{-J(\delta_1+\delta_2+\delta_3)} \Big/ \left(\sum_{j \in \mathcal{J}} e^{-J\delta_j}\right)^3 < 0.$$

3 Supplement for the Empirical Illustration

Table S4: Top 50 Brands

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

Elasticities for the Main Specifications. Tables S5 and S6 give the estimated average (over product types and markets) price elasticities of demands for the main specifications.

Table S5: Average Price Elasticities for the Three-Level NL Models

Type	3NL1				3NL2			
	Own	Cross			Own	Cross		
		Same subgroup	Same group	Different group		Same subgroup	Same group	Different group
1	-3.442	0.152	0.118	0.005	-3.440	0.177	0.131	0.007
2	-3.462	0.378	0.207	0.003	-3.547	0.316	0.085	0.004
3	-3.907	0.314	0.234	0.004	-3.975	0.234	0.125	0.006
4	-2.900	0.372	0.269	0.004	-3.034	0.244	0.103	0.005
5	-2.758	0.119	0.095	0.004	-2.776	0.116	0.084	0.006
6	-2.865	0.370	0.296	0.004	-3.156	0.194	0.094	0.006
7	-3.632	0.270	0.182	0.003	-3.714	0.196	0.077	0.005
8	-2.898	0.346	0.272	0.004	-3.008	0.185	0.086	0.006
9	-2.807	0.307	0.167	–	-2.026	1.106	–	0.003
10	-1.868	–	0.307	0.005	-1.488	–	0.624	0.007
11	-3.718	0.231	0.116	0.002	-3.503	0.468	0.313	0.003
12	-2.334	–	0.163	0.002	-2.139	–	0.286	0.003
13	-2.595	–	0.048	0.002	-2.333	–	0.234	0.003
14	-2.888	0.211	0.132	0.002	-2.709	0.440	0.333	0.004
15	-2.060	–	0.207	0.003	-1.842	–	0.360	0.004
16	-3.501	0.219	0.051	0.002	-2.723	1.019	0.790	0.003
17	-3.922	–	0.096	0.002	-3.186	–	0.717	0.003

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

Table S6: Average Price Elasticities for the IPDL Model

	Own	Cross																
Type		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-3.107	0.195	0.091	0.071	0.078	0.077	-0.026	-0.047	-0.039	0.004	0.010	-0.011	-0.003	0.108	-0.016	-0.008	0.064	-0.060
2	-3.203	0.059	0.323	0.064	0.068	-0.036	0.227	-0.032	-0.028	0.095	0.186	-0.073	-0.066	-0.008	-0.004	0.000	-0.006	-0.001
3	-3.560	0.068	0.093	0.334	0.084	-0.031	-0.006	0.235	-0.015	0.003	-0.055	0.186	-0.063	-0.110	0.157	-0.094	-0.051	0.216
4	-2.651	0.062	0.082	0.070	0.355	-0.037	-0.017	-0.028	0.256	0.003	-0.020	-0.032	0.249	-0.051	-0.042	0.242	-0.006	0.002
5	-2.513	0.062	-0.047	-0.027	-0.039	0.142	0.034	0.053	0.042	-0.006	-0.011	0.009	-0.003	0.085	-0.004	-0.016	0.050	-0.038
6	-2.581	-0.025	0.324	-0.006	-0.020	0.038	0.386	0.056	0.042	0.127	0.244	-0.086	-0.095	-0.013	0.005	-0.009	-0.009	0.009
7	-3.319	-0.035	-0.037	0.183	-0.028	0.047	0.046	0.266	0.055	-0.005	-0.063	0.158	-0.052	-0.090	0.129	-0.083	-0.043	0.176
8	-2.651	-0.032	-0.036	-0.013	0.259	0.043	0.038	0.061	0.334	-0.005	-0.036	-0.013	0.251	-0.055	-0.037	0.236	-0.006	0.012
9	-1.945	0.002	0.077	0.002	0.002	-0.004	0.072	-0.004	-0.003	0.912	0.060	-0.016	-0.014	0.005	0.005	0.006	0.005	0.005
10	-1.388	0.009	0.266	-0.051	-0.021	-0.013	0.244	-0.073	-0.043	0.107	–	0.489	0.501	0.036	-0.024	0.006	0.019	-0.040
11	-3.238	-0.005	-0.051	0.089	-0.020	0.004	-0.042	0.098	-0.011	-0.013	0.247	0.386	0.270	-0.029	0.065	-0.045	-0.012	0.082
12	-2.097	-0.002	-0.044	-0.029	0.146	-0.001	-0.044	-0.029	0.146	-0.011	0.233	0.248	–	-0.005	-0.032	0.143	0.011	-0.017
13	-2.379	0.043	-0.005	-0.047	-0.027	0.043	-0.006	-0.047	-0.027	0.004	0.015	-0.026	-0.006	–	0.199	0.219	0.036	-0.053
14	-2.492	-0.008	-0.003	0.086	-0.029	-0.003	0.002	0.092	-0.024	0.005	-0.015	0.075	-0.040	0.287	0.381	0.266	-0.013	0.081
15	-1.760	-0.006	0.001	-0.061	0.185	-0.012	-0.006	-0.067	0.179	0.007	0.004	-0.057	0.184	0.338	0.282	–	0.013	-0.041
16	-2.631	0.027	-0.004	-0.022	-0.003	0.027	-0.004	-0.022	-0.003	0.004	0.009	-0.009	0.010	0.040	-0.009	0.010	0.784	0.734
17	-3.287	-0.022	-0.001	0.084	0.001	-0.017	0.004	0.089	0.006	0.004	-0.018	0.067	-0.016	-0.049	0.057	-0.026	0.663	–

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

Bibliography

Allen, R. and J. Rehbeck (2019): "Hicksian Complementarity and Perturbed Utility Models," Available at SSRN: https://ssrn.com/abstract=2801345.

Anderson, S. P., A. De Palma, and J. F. Thisse (1988): "A representative consumer theory of the logit model," *International Economic Review*, 29, 461–466.

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

Berry, S., A. Eizenberg, and J. Waldfogel (2016): "Optimal product variety in radio markets," *The RAND Journal of Economics*, 47, 463–497.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

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

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 25, 242–262.

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

Björnerstedt, J. and F. Verboven (2016): "Does merger simulation work? Evidence from the Swedish analgesics market," *American Economic Journal: Applied Economics*, 8, 125–64.

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*, 28, S17–S44.

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7, 83–115.

Compiani, G. (2019): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," Unpublished.

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.

Deaton, A. and J. Muellbauer (1980): "An almost ideal demand system," *The American economic review*, 70, 312–326.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic competition and optimum product diversity," *The American Economic Review*, 67, 297–308.

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.

Ershov, D., J.-W. Laliberté, and S. Orr (2018): "Mergers in a Model with Complementarity," *Available at SSRN: https://ssrn.com/abstract=3295232*.

Fosgerau, M., D. McFadden, and M. Bierlaire (2013): "Choice probability generating functions," *Journal of Choice Modelling*, 8, 1–18.

Fosgerau, M., E. Melo, A. De Palma, and M. Shum (2018): "Discrete Choice and Rational Inattention: A General Equivalence Result," *Available at SSRN: https://ssrn.com/abstract=2889048 or http://dx.doi.org/10.2139/ssrn.2889048*.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

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

Galichon, A. and B. Salanié (2015): "Cupid's invisible hand: Social surplus and identification in matching models," Available at SSRN: https://ssrn.com/abstract=1804623.

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

Goldberg, P. K. (1995): "Product differentiation and oligopoly in international markets: The case of the US automobile industry," *Econometrica*, 63, 891–951.

Grigolon, L. (2018): "Blurred boundaries: a flexible approach for segmentation applied to the car market," Unpublished.

Hellerstein, R. (2008): "Who bears the cost of a change in the exchange rate? Pass-through accounting for the case of beer," *Journal of international economics*, 76, 14–32.

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

Iaria, A. and A. Wang (2019): "Identification and Estimation of Demand for Bundles: The Case of the RTE Cereal Industry," Unpublished.

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

Krugman, P. R. (1979): "Increasing returns, monopolistic competition, and international trade," *Journal of international Economics*, 9, 469–479.

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

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

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," Academic Press New York, 105–142.

——— (1978): "Modeling the choice of residential location," *Transportation Research Record*.

——— (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.

McFadden, D. L. and M. Fosgerau (2012): "A theory of the perturbed consumer with general budgets," NBER Working Paper No. 17953.

Michel, C. and S. Weiergraeber (2019): "Estimating Industry Conduct in Differentiated Products Markets," Unpublished.

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

Norets, A. and S. Takahashi (2013): "On the surjectivity of the mapping between utilities and choice probabilities," *Quantitative Economics*, 4, 149–155.

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

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

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

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

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

Simon, C. P. and L. Blume (1994): *Mathematics for economists*, New York, NY: W.W. Norton Company.

Small, K. A. (1987): "A discrete choice model for ordered alternatives," *Econometrica*, 55, 409–424.

Swait, J. and A. A. Marley (2013): "Probabilistic choice (models) as a result of balancing multiple goals," *Journal of Mathematical Psychology*, 57, 1–14.

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

Verboven, F. (1996a): "International price discrimination in the European car market," *The RAND Journal of Economics*, 27, 240–268.

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

Villas-Boas, S. (2007): "Vertical relationships between manufacturers and retailers: Inference with limited data," *The Review of Economic Studies*, 74, 625–652.

Chapter 2

The Flexible Inverse Logit Model

1 Introduction

Estimation of structural demand models for differentiated products plays an important role in economics. It allows to recover key parameters characterizing consumers' preferences so as to better understand their choices (e.g., by computing willingnesses to pay for product characteristics, price elasticities of demand, etc.). On top of that, it is the starting point of many economic questions of interest, including market power (Berry et al., 1995; Nevo, 2001), mergers (Nevo, 2000), products' entry (Petrin, 2002; Gentzkow, 2007), trade policy (Goldberg, 1995; Verboven, 1996; Berry et al., 1999), taxes (Griffith et al., 2019), cost pass-through (Miller et al., 2016), etc. As highlighted by the theoretical literature, the shape of the demand function, described by its slope and its curvature, drives the answers to these questions.[1] Then, assuming a supply model (e.g., price competition), the accuracy of the answers hinges on the ability of the demand model to yield general or unrestricted demand shape, i.e., on its ability to yield rich substitution patterns (roughly speaking, on its "flexibility"). Overall, this calls for a demand model which yields substitution patterns that are data-driven, rather than model-driven, and which ideally is easy and fast to estimate.

This paper pursues this goal by developing the flexible inverse logit (FIL) model. The FIL model is an inverse demand model for products that are differen-

[1] In monopolistically competitive and in oligopolistic models, the profit maximizing level of cost pass-through depends on both the slope (i.e., its first derivatives) and the curvature (i.e., its second derivatives) of the demand function (see Bulow and Pfleiderer, 1983; Weyl and Fabinger, 2013; Mrázová and Neary, 2017). For example, in monopolistically competitive markets, absolute pass-through is less than one if and only if log-demand is convex (Bulow and Pfleiderer, 1983; Weyl and Fabinger, 2013).

tiated in prices and in characteristics that may be observed or unobserved by the modeller. It accommodates rich substitution patterns, including complementarity, and is consistent with an underlying structural model of heterogeneous and utility-maximizing consumers. With aggregate data at hand (i.e., data on market shares, prices and product characteristics), the FIL model can be estimated by a linear instrumental variable (IV) regression, which allows to deal with the endogeneity of prices and market shares, due to the presence of structural error terms that represent unobserved (by the modeller, but observed by consumers and firms) product characteristics. The FIL model can be applied to various topics in industrial organization, international trade, and public economics; and it can be used to answer relevant policy questions, such as the effects of mergers, products' entry, and changes in regulation.[2]

The standard practice to structurally estimate demands for differentiated products based on aggregate data assumes an additive random utility model (ARUM). It then computes the associated demands that are estimated using the method developed by Berry (1994), which targets the inverse demands and thus involves an inversion from demands to their inverse that must be performed numerically when inverse demands are not closed form. Since Berry et al. (1995), the state-of-the-art approach is the BLP method, which assumes an RCL model and estimates it using a nested-fixed point algorithm.[3] The popularity of the BLP method is due to its ability to accommodate very rich substitution patterns while dealing with the endogeneity of prices and market shares.[4] However, the BLP approach faces some practical difficulties. Flexibility of the RCL model can be difficult to obtain in practice because it requires many random coefficients, which are not easily identified in applications; besides estimating an RCL model can be painful and time-consuming since it requires nonlinear optimization, numerical inver-

[2]In particular, it possesses the main features that make it appealing for merger evaluation purposes, as highlighted by Pinkse and Slade (2004). It imposes no specific restriction on the price elasticities; it is easily and fastly estimated by linear IV regression using standard computer softwares; and it can handle very large choice sets. To demonstrate its use for merger simulation, I provide, in Appendix E, a preliminary analysis of the Post-Nabisco merger in the ready-to-eat cereals industry that occurred in January 1993. Assuming that the merger had no effect on non-merging firm's prices, I find that the structural model of demand (FIL model) and supply (static oligopolistic price competition model) predicts pretty well the merging firms' price increase, as directly estimated using pre- and postmerger data.

[3]Since the RCL model yields inverse demands that are not closed form, Berry et al. (1995) propose to estimate demands by inverting them using a contraction mapping nested into a generalized method-of-moments (GMM) minimization procedure.

[4]McFadden and Train (2000) shows that any random utility model can be approximated by an RCL model.

sion and simulation of market shares.[5] Another widely used approach uses the nested logit models.[6] Because they have linear-in-parameters inverse demands, the nested logit models avoid the difficulties of the BLP approach by just requiring linear instrumental variable (IV) regressions (see Berry, 1994; Verboven, 1996). However, they have been criticized on the ground that they yield substitution patterns that are too restrictive and that they require the modeller to take a stand on the relevant dimensions along which nests can be defined.

The approach of this paper contrasts with the standard practice by specifying a linear-in-parameters inverse demand model, which just requires linear IV regression for estimation, and then by showing that it is consistent with a (utility) model of heterogeneous consumers.[7] Fosgerau et al. (2019) propose a general method to construct linear-in-parameters inverse demand models based on nesting, which yield richer substitution patterns than the nested logit model by allowing any nesting structure with overlapping and non-overlapping nests. However, this construction has two drawbacks with respect to the BLP approach: it requires the modeller to choose the nesting structure before estimation; and it yields substitutions patterns that do not depend on product characteristics directly – except those used to construct the nesting structure.[8]

[5]This implies dealing with their associated issues of local optima and choice of starting values, accuracy of the contraction mapping and of numerical integration (see e.g., Skrainka and Judd, 2011; Dubé et al., 2012; Knittel and Metaxoglou, 2014). See Conlon and Gortmaker (2018) for the current best practices in the estimation of structural demand models using BLP approach. Other approaches have been proposed. Dubé et al. (2012) transform the BLP's GMM minimization into a mathematical program with equilibrium constraints (MPEC), which minimizes the GMM objective function subject to the constraint that observed market shares be equal to predicted market shares. Lee and Seo (2015) approximate by linearization the nonlinear system of market shares for the RCL model and thus invert it analytically. Salanié and Wolak (2019) propose another approximation which leads to a linear IV regression.

[6]The nested logit models are commonly used by antitrust practitioners and competition authorities (e.g., the European Commission estimated nested logit models to simulate mergers for the Lagardère/Natexis/VUP (2004), TomTom/Tele Atlas (2008), Unilever/Sara Lee (2010) cases; see CCR - Competition Competence Report Autumn 2013/1) and by academics (see e.g., Björnerstedt and Verboven, 2016; Berry et al., 2016, for recent papers that estimate nested logit models with aggregate data).

[7]This paper adopts a fully parametric approach. By contrast, Compiani (2019) develops a non-parametric approach to estimate inverse demands in differentiated products markets based on aggregate data. His approach does not make any distributional assumptions on unobservables and imposes minimal functional form restrictions based on economic theory. However, he rules out complementarity, as defined by a negative cross-price derivative of demand. Using restrictions on inverse demands allows him to reduce the very large number of parameters to be estimated required in non-parametric settings. However, it still requires a large number of parameters and thus needs large enough datasets for estimation.

[8]In particular, Fosgerau et al. (2019) propose the inverse product differentiation logit (IPDL) model based on a predetermined segmentation of the market in which the multiple dimensions

The FIL model developed in the present paper overcomes these two drawbacks. Based on the construction proposed by Fosgerau et al. (2019), the FIL model uses a flexible nesting structure with a nest for each pair of products and an associated nesting parameter (see Chu, 1989; Koppelman and Wen, 2000; Davis and Schiraldi, 2014), which is consistent with a member of the class of model of heterogeneous and utility-maximizing consumers studied by Allen and Rehbeck (2019). Nesting in the FIL model is thus just a way to fully parametrize the matrix of price elasticities of demand and does not require the modeller to choose the nesting structure before estimation.[9] Then, the nesting parameters are projected into product characteristics space: in spirit to Pinkse et al. (2002), each parameter is defined as a function of the distance between products into the product characteristics space, known up to some parameters to be estimated. This projection imposes the same restrictions on the (inverse) demand function as the linear-in-characteristics ARUM – namely that only differences in characteristics determine the (inverse) demand, not the identity of products ; otherwise, it makes use of Bernstein polynomials for its shape to be data-driven.[10]

The FIL model is flexible in the sense of Diewert (1974) in a large class of inverse demand models imposing minimal restrictions on consumers' behavior,[11] except that it does not allow for income effect.[12] It is easily estimated by linear IV regression, which clarifies its empirical identification (in terms of instruments needed), and eases and accelerates its estimation (e.g., by application of the two-stage least squares estimator). The FIL model allows for complementarity in demand (i.e., negative cross-price derivatives of demand) and in utility (i.e., positive cross second-order partial derivatives of the utility function with respect to quantity). This is an important feature of the FIL model since many important economic questions hinge on the extent to which products are independent, substitutes or complements: in particular, it directly affects the incentive for firms to

of segmentation are allowed to cross in any way. The IPDL model extends the nested logit models to allow richer substitution patterns, including complementarity, but requires the modeller to choose the relevant dimensions of segmentation.

[9]This implies that its cross-price elasticities are not constrained by a predetermined segmentation of the market; however, it can still exploit product segmentation by adding segment fixed effects as product characteristics.

[10]Note that McFadden and Train (2000) use Bernstein polynomials to show the flexibility of the RCL model.

[11]This means that, observing a vector of market shares, it can match that vector of market shares as well as own- and cross-price elasticities.

[12]This is its main drawback. This is the case of any model derived from the construction based on nesting proposed by Fosgerau et al. (2019). Extending it to allow for income effect is beyond the scope of the present paper.

introduce a new product on the market (Gentzkow, 2007), to merge (Ershov et al., 2018), to bundle their products (Iaria and Wang, 2019), etc.[13] In this respect, the nesting parameters are interaction parameters, similar to those of Gentzkow (2007), determining the way products interact in utility and the type of relationship between them: products are substitutes (resp., complements) in utility, if their interaction parameter is positive (resp., negative).

Lastly, this paper studies the ability of the FIL model to provide rich substitution patterns. Using Monte Carlo simulations, it assesses its capacity to replicate the substitution patterns of the RCL model, which is used on the ground that it yields rich substitution patterns. They show that the FIL model provides a good approximation of the RCL model for specifications that are commonly used in the literature.

The remainder of the paper is organized as follows. Section 2 presents the general setting used in this paper. Section 3 introduces and studies the FIL model. Section 4 develops the methods to estimate it with data on market shares, prices and product characteristics. Section 5 shows, using Monte-Carlo simulations, the ability of the FIL model to replicate the substitution patterns of the RCL model for specifications that are commonly used in the literature. Section 6 concludes.

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

Let $\mathcal{J} = \{0,\ldots,J\}$. A vector $\mathbf{s} \in \mathbb{R}^{J+1}$ refers to $\mathbf{s} \equiv \left(s_0,\ldots,s_J\right)^\intercal \in \mathbb{R}^{J+1}$, and $\mathbf{s}_{-0} \equiv (s_1,\ldots,s_J)^\intercal$. $|\mathbf{s}|$ denotes the 1-norm of the vector $\mathbf{s}$, $|\mathbf{s}| = \sum_{j\in\mathcal{J}} |s_j|$.

Δ_J denotes the J-dimensional unit simplex: $\Delta_J \equiv \left\{\mathbf{s} \in [0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$, $\mathrm{int}\left(\Delta_J\right) \equiv \left\{\mathbf{s} \in (0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$ is its interior, and $\mathrm{bd}(\Delta_J) \equiv \Delta_J \backslash \mathrm{int}\left(\Delta_J\right)$ is its boundary.

Let $\mathbf{G} = \left(G_0,\ldots,G_J\right) : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a vector function composed of functions $G_j : \mathbb{R}^{J+1} \to \mathbb{R}$. The Jacobian matrix $\mathbf{J}^{\mathbf{s}}_{\mathbf{G}}(\bar{\mathbf{s}})$ of $\mathbf{G}$ with respect to $\mathbf{s}$ at $\bar{\mathbf{s}}$ has entries ij given by $\frac{\partial G_i(\bar{\mathbf{s}})}{\partial s_j}$. The matrix $\left[\mathbf{J}^{\mathbf{s}}_{\mathbf{G}}(\bar{\mathbf{s}})\right]_0$ corresponds to the matrix $\mathbf{J}^{\mathbf{s}}_{\mathbf{G}}(\bar{\mathbf{s}})$ after removing its first row and its first column.

A univariate function $\mathbb{R} \to \mathbb{R}$ applied to a vector is a coordinate-wise application of the function, e.g., $\ln(\mathbf{s}) = \left(\ln(s_0),\ldots,\ln\left(s_J\right)\right)$. $\mathbf{0}_J = (0,\ldots,0)^\intercal \in \mathbb{R}^J$ and $\mathbf{1}_J = (1,\ldots,1)^\intercal \in \mathbb{R}^J$ are vectors consisting of zeroes and ones, respectively.

[13] For example, a merger between firms selling complements reduces prices (Cournot, 1838).

$\mathbf{I}_{J,J} \in \mathbb{R}^{J \times J}$ is the $J \times J$ identity matrix and $\mathbf{1}_{J,J} \in \mathbb{R}^{J \times J}$ is the $J \times J$ matrix consisting of ones.

2 Setting

This section first presents the general setting used in this paper. It gives the main ingredients required to identify and estimate demand models for differentiated products using aggregate data. In particular, it highlights that it is the inverse demand that is targeted during estimation. Then, it introduces the general method developed by Fosgerau et al. (2019) to construct inverse demand models based on nesting, which is used in Section 3 to build the flexible inverse logit model.

2.1 General Setting

Consider a population of consumers choosing from a choice set $\mathcal{J} \equiv \{0, \ldots, J\}$ of $J+1$ differentiated products, where products $j \in \mathcal{J}_0 \equiv \{1, \ldots, J\}$ are the inside products and product $j = 0$ is the outside good. Let $p_j \in \mathbb{R}$ be the price of product j and $\mathbf{x}_j \in \mathbb{R}^K$ be the vector of K observed characteristics of product j, with $\mathbf{p} \equiv \left(p_0, \ldots, p_J\right)$ and $\mathbf{x} \equiv \left(\mathbf{x}_0, \ldots, \mathbf{x}_J\right)$.

Following the prevailing literature (Berry, 1994; Berry et al., 1995), let $\xi_j \in \mathbb{R}$ be the j-product unobserved characteristics term, with $\boldsymbol{\xi} \equiv \left(\xi_0, \ldots, \xi_J\right)$. The vector $\boldsymbol{\xi}$ contains the structural error terms of the demand model, which are considered to be observed by consumers and firms but not by the modeller.

Consider the demand system

$$\boldsymbol{\sigma} = \left(\sigma_0, \ldots, \sigma_J\right) : \mathcal{X} \to \operatorname{int}\left(\Delta_J\right),$$

where $\mathcal{X}$ is the support of $(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi})$ and where the function $\boldsymbol{\sigma}$ gives the vector $\mathbf{s} = \left(s_0, \ldots, s_J\right) \in \operatorname{int}\left(\Delta_J\right)$ of nonzero observed market shares, that is, it yields the system of market shares which equates the vector $\mathbf{s}$ of observed market shares to the vector $\boldsymbol{\sigma}$ of predicted (by the model) market shares,

$$\mathbf{s} = \boldsymbol{\sigma}(\mathbf{p}, \mathbf{x}, \boldsymbol{\xi}; \boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\theta}$ denotes the vector of structural parameters to be estimated. The structural parameters are the key parameters describing consumers' preferences, which are invariant to changes in economic policy, such as taxes, or in firms' strategy, such as pricing strategies, product characteristics, new products (see Hurwicz, 1966).

The demand system (1) assumes that the terms ξ_j, $j \in \mathcal{J}$, are scalars, that there is no income effect (since $\boldsymbol{\sigma}$ is independent of income), and that demands are positive (i.e., $\sigma_j > 0$ for all $j \in \mathcal{J}$).

Index restriction Following Berry and Haile (2014), I further assume an index restriction. Specifically, I partition the set of K product characteristics as $\mathbf{x} = \left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right)$, where $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are K_1 linear product characteristics and $K_2 = K - K_1$ nonlinear product characteristics, respectively. Then, I define linear indexes as

$$\delta_j \equiv \mathbf{x}_j^{(1)}\boldsymbol{\beta} - \alpha p_j + \xi_j, \quad j \in \mathcal{J}, \tag{2}$$

where $\alpha > 0$ is the consumers' price sensitivity (i.e., their marginal utility of income) and where $\boldsymbol{\beta}$ captures the consumers' taste for characteristics $\mathbf{x}^{(1)}$. The vector $(\alpha, \boldsymbol{\beta})$ is often referred to as the vector of linear parameters. I normalize the indexes of each inside product relative to that of the outside good by setting $\delta_0 = 0$, so that $\boldsymbol{\delta}$ belongs to $\mathbb{R}_0^{J+1} \equiv \{\boldsymbol{\delta} : \delta_0 = 0\}$. This normalization is required for identification (see Proposition 1 for details).

This means that the system of market shares (1) can be rewritten as

$$\mathbf{s} = \boldsymbol{\sigma}\left(\boldsymbol{\delta}, \mathbf{x}^{(2)}; \boldsymbol{\mu}\right), \tag{3}$$

where $\boldsymbol{\delta} \equiv \left(\delta_0, \ldots, \delta_J\right)$ and $\boldsymbol{\mu}$ is often referred to as the vector of nonlinear parameters.

The index restriction implies that $\mathbf{x}^{(1)}$, $\mathbf{p}$ and $\boldsymbol{\xi}$ enter demands only through the product indexes $\boldsymbol{\delta}$. By constrast, $\mathbf{x}^{(2)}$ can enter demands in an unrestrictive way. In the logit model, there is no $\mathbf{x}^{(2)}$. By contrast, in the RCL model, $\mathbf{x}^{(2)}$ are the characteristics that have a random coefficient. Demands (3) depending on $\mathbf{x}^{(2)}$ make the own- and cross-price elasticities depending on product characteristics directly; otherwise, they do not depend on product characteristics directly, except when parameters depend on product characteristics - as it is, for example, the case of the nested logit model.

Inverse Demand Invertibility in $\boldsymbol{\delta}$ of the system of market shares (3) is crucial for identification and estimation. Berry et al. (2013) show that their "connected substitutes" structure is sufficient for invertibility.[14] In Monardo (2019), I provide a new result of invertibility for demands that accommodate some substitution patterns that are not allowed by Berry et al. (2013), including complementarity as defined by a negative cross-price derivative of demand; this result applies to the FIL model developed in Section 3.

If $\boldsymbol{\sigma}$ is invertible in $\boldsymbol{\delta}$, there exists an inverse demand function $\boldsymbol{\sigma}^{-1}$ given by

$$\delta_j = \sigma_j^{-1}\left(\mathbf{s}, \mathbf{x}^{(2)}; \boldsymbol{\mu}\right), \quad j \in \mathcal{J},$$

so that the structural error terms ξ_j can be written as

$$\xi_j(\boldsymbol{\theta}) = \sigma_j^{-1}\left(\mathbf{s}, \mathbf{x}^{(2)}; \boldsymbol{\mu}\right) - \mathbf{x}_j^{(1)}\boldsymbol{\beta} + \alpha p_j, \quad j \in \mathcal{J}. \tag{4}$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\mu})$.

Equation (4), which gives each structural error term as a function of the data (i.e., market shares, prices and product characteristics) and parameters $\boldsymbol{\theta}$ to be estimated, shows that the inverse demand function, not the demand function itself, is the target of estimation.

Instruments Following the prevailing literature (Berry, 1994; Berry et al., 1995; Nevo, 2001), I assume that product characteristics $\mathbf{x}$ are exogenous, i.e., they are uncorrelated with the structural error terms. However, prices and market shares in the right-hand side of Equation (4) are likely to be endogenous. Prices are endogenous because, as it is typically assumed in price competition models with differentiated products, firms consider both observed and unobserved product characteristics when they set their prices. Market shares are endogenous since they are determined by the system of equations (1), where each market share depends on the entire vectors of endogenous prices and of unobserved product characteristics, and because consumers choose products while considering unobserved product characteristics.

Then, provided that there exists appropriate instruments $\mathbf{z}$ for prices and market shares, following Berry (1994), one can estimate demands (3) based on the

[14] The connected substitutes structure requires that (i) products be weak substitutes, i.e., everything else equal, an increase in δ_j weakly decreases demand σ_i for all other products; and (ii) the "connected strict substitution" condition hold, i.e., there is sufficient strict substitution between products to treat them in one demand system.

following conditional moment restrictions

$$\mathbb{E}\left[\xi_j(\boldsymbol{\theta})|\mathbf{z}\right] = 0, \quad j \in \mathcal{J}, \tag{5}$$

where $\xi_j(\boldsymbol{\theta})$ is given by Equation (4).

Both cost-based instruments and BLP instruments can be used (see e.g., Berry and Haile, 2014; Reynaert and Verboven, 2014; Armstrong, 2016; Gandhi and Houde, 2017, etc.). Cost-based instruments separate exogenous variation in prices due to exogenous changes in costs from endogenous variation in prices due to changes in unobserved product characteristics. They are valid under the assumption that variation in cost shifters is correlated with variation in prices, but not with changes in unobservable product characteristics. However, they may not be good instruments for market shares, since costs do not directly shift the endogenous market shares but instead only affect them through prices.[15]

BLP instruments are functions of the characteristics of competing products and are valid under the assumption that product characteristics $\mathbf{x}$ are exogenous. They separate exogenous variation in prices due to changes in $\mathbf{x}$ from endogenous variation in prices due to changes in unobserved product characteristics. They are commonly used to instrument prices with the idea that characteristics of competing products are correlated with prices since the (equilibrium) markup of each product depends on how close products are into product characteristics space (products with close substitutes tend to have low markups and thus low prices relative to costs). They are also appropriate to instrument market shares. BLP instruments can suffice for (theoretical) identification, but may be weak in practice, thereby making cost-based instruments useful (see e.g., Reynaert and Verboven, 2014).[16]

2.2 Linear-in-Parameters Inverse Demand Models

Equations (4) and (5), when combined with parametric functional form restrictions, serve as a basis for demand estimation. When inverse demands σ_j^{-1} have a closed form, one can resort to IV regression techniques for estimation. If, in addition, they are linear in parameters $\boldsymbol{\mu}$, then estimation just requires linear IV

[15]The weak instruments problem occurs when instruments are only weakly correlated with the endogenous variables. See Andrews et al. (2018) on how to test for weak instruments in applications.

[16]Note that Armstrong (2016) discusses the strength of the BLP instruments as the number of products increases.

regressions.

Fosgerau et al. (2019) show how to construct linear-in-parameter inverse demand models based on nesting, thereby generalizing the nested logit models by allowing any nesting structure, with overlapping or non-overlapping nests, while just requiring linear IV regressions for estimation.[17] Inverse demands of such models are of the form of

$$\sigma_j^{-1}(\mathbf{s};\boldsymbol{\mu}) \equiv \ln G_j(\mathbf{s};\boldsymbol{\mu}) + c = \delta_j, \quad j \in \mathcal{J}, \tag{6}$$

where δ_j is given by Equation (2) and $c \in \mathbb{R}$ is a market-specific constant, and where $\ln G_j$ is defined by

$$\ln G_j(\mathbf{s};\boldsymbol{\mu}) \equiv \mu_j \ln\left(s_j\right) + \sum_{g \in \mathcal{G}(j)} \mu_g \ln\left(\sum_{k \in g} s_k\right), \quad j \in \mathcal{J}, \tag{7}$$

where $\mathcal{G}(j)$ is the set of nests containing product j, and where $\boldsymbol{\mu} \equiv \left((\mu_j)_{j\in\mathcal{J}}, (\mu_g)_{g\in\mathcal{G}}\right)$, with $\mathcal{G}$ the finite set of all nests, is such that $\mu_j + \sum_{g\in\mathcal{G}(j)} \mu_g = 1$, for all $j \in \mathcal{J}$, with $\mu_j > 0$ for all $j \in \mathcal{J}$ and $\mu_g \geq 0$ for all $g \in \mathcal{G}$.

The logit and the nested logit models are the simplest special cases (see e.g., Berry, 1994; Verboven, 1996). For the logit model,

$$\ln G_j(\mathbf{s}) \equiv \ln(s_j). \tag{8}$$

Partition the choice set $\mathcal{J}$ into nests, and further partition each nest into subnests. Then, one obtains the three-level nested logit model with

$$\ln G_j(\mathbf{s};\mu_1,\mu_2) \equiv \left(1 - \sum_{d=1}^{2} \mu_d\right) \ln\left(s_j\right) + \mu_1 \ln\left(\sum_{k\in g} s_k\right) + \mu_2 \ln\left(\sum_{k\in g|h} s_k\right), \tag{9}$$

for product $j \in \mathcal{J}_0$ in nest g and subnest $h|g$, where $\mu_1, \mu_2 \geq 0$ with $\mu_1 + \mu_2 < 1$. Indeed, setting $\gamma_1 = \mu_1 + \mu_2$ and $\gamma_2 = \mu_1$, one obtains Equation (10) of Verboven (1996), with the constraints $0 \leq \gamma_2 \leq \gamma_1 < 1$ that make it consistent with random utility maximization. Note that, for the nested logit models, the inverse demand of the outside good is given by

$$\ln G_0(\mathbf{s}) \equiv \ln(s_0),$$

[17] See Corollary A in the supplement of Fosgerau et al. (2019).

since the outside good is assumed to be the only member of its nest.

The construction proposed in Equations (6) and (7) leads to models that have several attractive features. They generalize the nested logit models by allowing for richer substitution patterns, and in particular complementarity. They require linear IV regressions for estimation, which clarifies their empirical identification in terms of instruments needed, and eases and accelerates their estimation by application of the two-stage least squares estimator. They can be sufficiently parsimonious (i.e., their number of parameters does not grow with the number of products), that they can fastly handle large choice sets.

However, this construction has two drawbacks with respect to the BLP approach. First, it requires the modeller to choose a nesting structure before estimation. For example, Fosgerau et al. (2019) build the inverse product differentiation logit (IPDL) model which, similarly to the PDL model of Bresnahan et al. (1997), generalizes the nested logit models by allowing an arbitrary, non-hierarchical nesting structure in which dimensions of segmentation are allowed to cross in any way. The IPDL model allows for richer substitution patterns than the nested logit models, including complementarity; however, it still requires the modeller to choose the relevant dimensions of segmentation.

The choice of the nesting structure can be problematic in applications. Consider for example the market for cars, where cars are assumed to belong to five segments: subcompact, compact, standard, intermediate, and luxury. Grigolon (2018) suggests a natural ordering of cars from subcompact to luxury, while Brenkers and Verboven (2006) consider a nested structure without prior ordering. Determining which of the two nesting structures best describes the market is not obvious.

Second, it leads to models with substitution patterns that depend on product characteristics only indirectly through the market shares, except for those used to construct the nesting structure.

3 The Flexible Inverse Logit Model

This section develops the flexible inverse logit (FIL) model, which overcomes the two drawbacks mentioned above. It is a flexible model that uses the construction of Equations (6) and (7) to build a flexible nesting structure with a nest for each pair of products and with an associated nesting parameter that drives the substitutions between these two products. Thus, in the FIL model, nesting does not require the modeller to identify relevant dimensions of segmentation.

Then, to make the substitution patterns depend on product characteristics directly, I project the nesting parameters into product characteristics space. In particular, in spirit to Pinkse et al. (2002), I replace each nesting parameter with a function of a measure of distance between products i and j into the product characteristics space formed by $\mathbf{x}^{(2)}$.

3.1 Specification

The FIL model uses a flexible nesting structure with a nest for each pair of products $(i,j) \in \mathcal{J}_0$, $i \neq j$, and with an associated parameter μ_{ij} for each nest. Its inverse demands are given by

$$\sigma_j^{-1}(\mathbf{s};\boldsymbol{\mu}) \equiv \ln G_j(\mathbf{s};\boldsymbol{\mu}) + c = \delta_j, \quad j \in \mathcal{J}_0, \tag{10}$$

where δ_j is given by Equation (2), and where $\ln G_j$ is defined by

$$\ln G_j(\mathbf{s};\boldsymbol{\mu}) \equiv \mu_j \ln\left(s_j\right) + \sum_{i \neq j} \mu_{ij} \ln\left(s_i + s_j\right), \quad j \in \mathcal{J}_0, \tag{11}$$

where $\boldsymbol{\mu} \equiv \left(\mu_1, \ldots, \mu_J, \mu_{11}, \ldots, \mu_{J-1,J}\right)$ satisfies the following assumption.

Assumption 1. The vector $\boldsymbol{\mu}$ satisfies the following constraints.

(i) $\mu_j + \sum_{i \neq j} \mu_{ij} = 1$ for all $j \in \mathcal{J}_0$

(ii) $\mu_j > 0$ for all $j \in \mathcal{J}_0$

(iii) $\mu_{ij} = \mu_{ji}$ for all $(i,j) \in \mathcal{J}_0^2$, $i \neq j$.

In addition, the outside good is the only member of its nest, then

$$\ln G_0(\mathbf{s};\boldsymbol{\mu}) = \ln(s_0), \tag{12}$$

and

$$\sigma_0^{-1}(\mathbf{s};\boldsymbol{\mu}) \equiv \ln(s_0) + c = 0, \tag{13}$$

where I recall that normalization $\delta_0 = 0$ is used.

Assumption 1 has an economic content and can be expressed in terms of demand shape restrictions. The FIL model yields positive demands that sum to one, that are invariant to translation in $\boldsymbol{\delta}$ (i.e, only differences in $\boldsymbol{\delta}$ determine demands, not their absolute value), that satisfy Slutsky symmetry and positive definiteness (i.e., the matrix of demand derivatives with respect to $\boldsymbol{\delta}$ is symmetric and positive definite) and the boundary condition that all the consumers choose a product $j \in \mathcal{J}$ when that product becomes infinitely attractive (i.e., when δ_j goes to infinity), while the others remain finitely attractive.

Several remarks are in order. First, Assumption 1 allows the nesting parameters μ_{ij} to be positive and negative. Then, the FIL model nests the logit model. It boils down to the logit model when $\mu_{ij} = 0$ for all $i, j \in \mathcal{J}_0$, $i \neq j$. As a consequence, the independence from irrelevance alternatives (IIA) property could be tested using standard Wald tests.[18] Lastly, one could also test (rather than impose) Slutsky symmetry by testing whether or not Assumption (iii) holds. Note, however, that Slutsky symmetry is key for rationalizability of the demand function, i.e., for it to be consistent with utility maximization. Then, if Slutsky symmetry is rejected by the data, any study based on the demand estimates (e.g., merger simulation) would not rely on an underlying utility model.

Demand Invertibility The following proposition establishes invertibility of the demand.

Proposition 1. Let Assumption 1 hold. Consider any vector $\mathbf{s} \in \operatorname{int}\left(\Delta_J\right)$ of nonzero market shares. Then, there exists a unique $\boldsymbol{\delta} \in \mathbb{R}_0^{J+1}$ such that

$$\boldsymbol{\delta} = \boldsymbol{\sigma}^{-1}(\mathbf{s}; \boldsymbol{\mu}) \Leftrightarrow \mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta}; \boldsymbol{\mu}),$$

where $\boldsymbol{\sigma}^{-1}$ is given by Equations (10) to (13).

Proof. See Appendix B.1. □

Equations (10) to (13) describe the inverse demand of the FIL model, i.e., its mapping from market shares $\mathbf{s}$ to product indexes $\boldsymbol{\delta}$. Proposition 1 establishes existence and uniqueness of the inverse mapping from product indexes to market shares (i.e., the demand), up to normalization. It states that product indexes $\boldsymbol{\delta}$ are identified up to an additive constant c from the vector of observed nonzero market shares $\mathbf{s}$ by the relations (10) to (13), where c is fixed by normalizing δ_0 to zero. The need for normalization is due to the feature of the FIL model whereby its demand function satisfies translation invariance in product indexes $\boldsymbol{\delta}$.

[18] However, the FIL model and the nested logit model are non-nested.

Microfoundation The FIL model is consistent with a consumer choosing a vector $\mathbf{s} \in \Delta_J$ of market shares of the differentiated products so as to maximize her quasi-linear utility given by[19]

$$u(\mathbf{s}) = \sum_{j \in \mathcal{J}} \delta_j s_j - \sum_{j \in \mathcal{J}} s_j \ln\left(s_j\right) + \sum_{j \in \mathcal{J}_0} \sum_{i \neq j} \mu_{ij} s_j \ln\left(\frac{s_j}{s_i + s_j}\right),$$

if $\mathbf{s} \in \Delta_J$ and $u(\mathbf{s}) = -\infty$ if $\mathbf{s} \notin \Delta_J$, where δ_j is given by Equation (2).

The FIL model can thus be seen as a representative consumer model (Fosgerau et al., 2019) or a standard continuous-choice model. It is worthwhile to mention that this form of utility can be derived, after an aggregation across consumers, from the utility model of heterogeneous and utility-maximizing consumers, studied by Allen and Rehbeck (2019). Parameters μ_{ij} thus parametrize the distribution of taste in the population of consumers.

Complements and Substitutes The parameters μ_{ij} of the FIL model relate to two different (but related) definitions of complementarity used in the empirical literature (see Appendix A.4). In the FIL model, the sign of the parameter μ_{ij} determines whether products i and j are complements or substitutes in utility. Indeed,

$$\frac{\partial^2 u(\mathbf{s})}{\partial s_i \partial s_j} = -\frac{\mu_{ij}}{s_i + s_j},$$

so that products i and j are substitutes in utility if $\mu_{ij} > 0$, complements in utility if $\mu_{ij} < 0$, and independent in utility if $\mu_{ij} = 0$.

The nesting parameters are thus parameters, similar to those of Gentzkow (2007), determining the way products interact in utility and thus the type of relationship between them. Gentzkow (2007) uses a discrete analogue of this definition in an ARUM based on bundles ij with utilities

$$u_{ij} = u_i + u_j + \Gamma_{ij},$$

so that products i and j are complements in utility if $\Gamma_{ij} > 0$, substitutes in utility if $\Gamma_{ij} < 0$ and independent in utility if $\Gamma_{ij} = 0$.

Each parameter μ_{ij} drives the cross-price elasticities between products i and

[19] Note that Assumption (i) has been used.

j. To see this, let $J = 3$ with $s_j = 1/4$ and $p_j = p$. Then,

$$\frac{\partial s_1}{\partial p_3}\frac{p_3}{s_1} > \frac{\partial s_1}{\partial p_2}\frac{p_2}{s_1} \Leftrightarrow \mu_{13} > \mu_{12}.$$

However, as suggested by preliminary simulations, there is no relationship between the sign of μ_{ij} and the sign of the cross-price derivative (or elasticity) of demand between products i and j. Rather, whether or not products i and j are complements or substitutes depends on the relation of these two products to the other products, as already highlighted by the theoretical literature (see e.g., Samuelson, 1974; Ogaki, 1990). In particular, these simulations show that the FIL model rules out complementarity when $\mu_{ij} \geq 0$ for all $i, j \in \mathcal{J}_0$ and $\mu_j > 0.5$ for all $j \in \mathcal{J}_0$.[20]

Flexibility The FIL model can be shown to be flexible in the sense of Diewert (1974) in the class of inverse demand models of the form of Equation (10), where $\mathbf{G}$ is homogeneous of degree one, with $G_0(\mathbf{s}) = s_0$, and where the Jacobian $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ is positive definite and symmetric on $\text{int}(\Delta_J)$.[21]

A demand system is said to be flexible in the sense of Diewert (1974) if it is able to provide a first-order approximation to any theoretically grounded demand system at a point in price space.[22] Equivalently, flexibility can also be viewed as the ability of the (direct or indirect) utility function to provide second-order approximations to any utility function. This is because the partial derivatives of the demand function can be uniquely derived from the second partial derivatives of the utility function.

Flexibility of the FIL model means that, observing market shares $\mathbf{s}^*$, the FIL can match that vector of market shares $\mathbf{s}^*$ as well as the true matrix of price derivatives of demand for inside products $\left[\mathbf{J}^{\mathbf{p}*}_{\sigma}\right]_0$. That is, if the FIL model is flexible, then one can choose $\boldsymbol{\delta}^*$, α^* and $\boldsymbol{\mu}^*$ such that $\mathbf{s}^* = \boldsymbol{\sigma}(\boldsymbol{\delta}^*; \boldsymbol{\mu}^*)$ and $\left[\mathbf{J}^{\mathbf{p}*}_{\sigma}\right]_0 = \left[\mathbf{J}^{\mathbf{p}}_{\sigma}(\boldsymbol{\delta}^*; \boldsymbol{\mu}^*)\right]_0$.

Flexibility of the FIL model is best understood by focusing on its inverse demand function and its corresponding matrix of derivatives. Indeed, flexibility

[20]Further simulations will be performed later.

[21]This is the class of generalized inverse logit (GIL) models developed by Fosgerau et al. (2019) where it is further assumed that $\ln G_0 = \ln(s_0)$, as it is the case of the logit, nested logit, and IPDL models.

[22]See Appendix A.2 for more details.

can be obtained by matching

$$\sigma_j^{-1}(\mathbf{s}^*;\boldsymbol{\mu}^*) = \ln G_j(\mathbf{s}^*;\boldsymbol{\mu}^*) + c = \delta_j^*, \quad j \in \mathcal{J},$$

and[23]

$$\left[\mathbf{J}_{\sigma}^{\mathbf{p}}(\delta^*;\boldsymbol{\mu}^*)\right]_0^{-1} = -\frac{1}{\alpha^*}\left[\frac{\mathbf{1}_{J,J}}{s_0^*} + [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s}^*;\boldsymbol{\mu}^*)]_0\right] = \left[\mathbf{J}_{\sigma}^{\mathbf{p}*}\right]_0^{-1}, \tag{14}$$

where $\ln G_j$ is given by Equations (11) and (12) so that $\left[\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s};\boldsymbol{\mu})\right]_0$ is given by

$$\left[\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s};\boldsymbol{\mu})\right]_0 = \begin{bmatrix} \frac{\mu_1}{s_1} + \sum_{i\neq 1}\frac{\mu_{i1}}{s_i+s_1} & \frac{\mu_{12}}{s_1+s_2} & \cdots & \frac{\mu_{1J}}{s_1+s_J} \\ \frac{\mu_{12}}{s_1+s_2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\mu_{1J}}{s_1+s_J} & \cdots & \cdots & \frac{\mu_J}{s_J} + \sum_{i\neq J}\frac{\mu_{iJ}}{s_i+s_J} \end{bmatrix}.$$

The FIL model yields flexible substitution patterns in the sense that it has a sufficient number of parameters to fully parametrize $\left[\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s};\boldsymbol{\mu})\right]_0$. At the opposite, the logit model fully sparsifies $\left[\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})\right]_0$,

$$\left[\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})\right]_0 = \begin{bmatrix} \frac{1}{s_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{s_J} \end{bmatrix},$$

which does not depend on parameters $\boldsymbol{\mu}$.

Flexibility of the FIL model is obtained as follows. One can match the off-diagonal entries ij of $\left[\mathbf{J}_{\sigma}^{\mathbf{p}*}\right]_0^{-1}$, by appropriately choosing $\mu_{ij} = \mu_{ij}^*$. Its diagonal entries are then automatically matched since the FIL model satisfies the Euler-type equation $\sum_{k\in\mathcal{J}}\frac{\partial \ln G_j(\mathbf{s}^*;\boldsymbol{\mu}^*)}{\partial s_k}s_k^* = 1$ for all $j \in \mathcal{J}$ (see Fosgerau et al., 2019). One can match the vector of market shares $\mathbf{s}^*$ by choosing the unique vector $\delta^* \in \mathbb{R}_0^{J+1}$ that is able to do so. The parameter α^* can be chosen so that $\mu_j > 0$ for all $j \in \mathcal{J}_0$.

The following proposition summarizes this discussion.

Proposition 2. The FIL model is flexible in the sense of Diewert (1974) in the class of inverse demand models of the form of Equation (10), where $\mathbf{G}$ is homogeneous of degree one, with $G_0 = s_0$, and where the Jacobian $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}$ is positive definite and symmetric on $\mathrm{int}(\Delta_J)$.

[23] See Appendix B.2 for details of computations.

Proof. See Appendix B.3. □

3.2 Projection into Product Characteristics Space

The FIL model yields substitution patterns that are flexible (in the sense of Diewert (1974)), but that depend on product characteristics only indirectly (through the market shares). I address this drawback by projecting the nesting parameters μ_{ij} into the product characteristics space formed by $\mathbf{x}^{(2)}$, based on the idea that the distance between products into this space drives substitution (closer products into this space should be more substitutable).

In spirit to Pinkse et al. (2002), I replace μ_{ij} with a function of a measure of distance $\mathbf{d}_{ij}^{(2)} \equiv (d_{ij1}^{(2)}, \ldots, d_{ijK_2}^{(2)})$ between products i and j into the product characteristics space formed by $\mathbf{x}^{(2)}$,[24]

$$\mu_{ij} = \mu\left(\mathbf{d}_{ij}^{(2)}; \gamma\right),$$

known up to some parameters γ to be estimated.

This projection makes the substitution patterns depending on characteristics $\mathbf{x}^{(2)}$. It maps the FIL model from product space into product characteristics space; thereby making it appealing to perform counterfactual exercises, such as the simulation or evaluation of mergers, products' entry, or changes in regulation. It also makes the price elasticities functions of a small number of parameters γ to be estimated.

With this transformation, the parameters μ_{ij} are no longer structural, since they are not invariant to changes in product characteristics by firms. However, the vector of parameters contained in the vector γ parametrizing the function μ are structural; they are similar in spirit to the standard deviations of the random coefficients in a RCL model, as they control for the distribution of valuation for product characteristics $\mathbf{x}^{(2)}$ in the population of consumers.

The relationship to other approaches, and in particular, with the ideal point (or address) approach (Anderson et al., 1992), where in this case the ideal points just consists of the locations of products into the product characteristics space

[24]This strategy has been successfully applied by Pinkse and Slade (2004) Slade (2004), Rojas (2008) for demand estimation purposes. See also Pinkse and Slade (1998). Note that I do not implement the semi-parametric estimator of Pinkse et al. (2002). In my model, their method would use a series expansion to approximate μ, and in turn, this would introduce an additional source of endogeneity. Indeed, in addition to the structural error ξ, their method adds an approximation error, due neglected expansion errors, that is a function of characteristics $\mathbf{x}^{(2)}$.

formed by $\mathbf{x}^{(2)}$, or with discrete-continuous models (e.g., Hanemann, 1984) will be studied later.

Economic Restrictions The FIL model projected into product characteristics space satisfies the same shape restrictions on the inverse demands as those that are imposed in the linear-in-characteristics ARUM, including the RCL model (Gandhi and Houde, 2017).

Observe first that the FIL model projected into product characteristics space has inverse demands given by

$$\sigma_j^{-1}\left(\mathbf{s},\mathbf{x}^{(2)};\gamma\right) = \ln G\left(s_j,\{s_k,\mathbf{d}_{kj}^{(2)}\}_{k\neq j};\gamma\right) + C, \tag{15}$$

where

$$\ln G\left(s_j,\{s_k,\mathbf{d}_{kj}^{(2)}\}_{k\neq j};\gamma\right) \equiv \ln\left(s_j\right) - \sum_{i\neq j}\mu\left(\mathbf{d}_{ij}^{(2)};\gamma\right)\ln\left(\frac{s_j}{s_i+s_j}\right), \tag{16}$$

with $C \in \mathbb{R}$ a market-specific constant (see Appendix C for more details).

Now, let $y_j \equiv (s_j,\mathbf{x}_j^{(2)})$ and $\mathbf{y}_{-j} \equiv (y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_J)$. Then, the inverse demands (15) can be rewritten in terms of y_j as $\sigma_j^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right)$. One can easily show that the FIL model projected into product characteristics space yields inverse demands which are symmetric, that is,

$$\sigma_j^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right) = \sigma_k^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right), \quad j\neq k,$$

so that one can set $\sigma_j^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right) = \sigma^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right)$; which are anonymous, that is,

$$\sigma^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right) = \sigma^{-1}\left(y_j,\mathbf{y}_{\rho(-j)};\gamma\right),$$

where $\rho(-j)$ is any permutation of the product indexes $-j$; and which are invariant to translation in $\mathbf{x}^{(2)}$, that is,

$$\sigma^{-1}\left(y_j+(0,c),\mathbf{y}_{-j}+(\mathbf{0},c\mathbf{1});\gamma\right) = \sigma^{-1}\left(y_j,\mathbf{y}_{-j};\gamma\right).$$

Overall, these shape restrictions mean that that the identity of products does not matter, only the differences in characteristics do.[25] The following proposition

[25]This approach builds on Doraszelski and Pakes (2007) who use symmetry and anonymity to reduce the dimensionality of value functions in the context of dynamic games, on Compiani (2019) who uses anonymity to reduce the number of parameters of the inverse demand to be

gives the properties of the FIL model projected into product characteristics space.

Proposition 3. The FIL model projected into product characteristics space has inverse demands given by Equations (15) and (16), which is symmetric, anonymous and invariant to translation in product characteristics.

Proof. See Appendix C. □

Measures of distance The measures of distance $\mathbf{d}_{ij}^{(2)}$ can be continuous or discrete. The paper focuses on continuous measures and uses the absolute value $\mathbf{d}_{ij}^{(2)} = |\mathbf{x}_i^{(2)} - \mathbf{x}_j^{(2)}|$ as measure.[26] Other measures, such as the euclidian distance, can also be used.

The projection can also exploit product segmentation, as the nested logit or the IPDL models do, by using categorical variables that group products into different segments. In this case, the measure of distance is discrete: for a categorical variable k, one can set $d_{ijk}^{(2)} = 1$ if products i and j belong to the same nest according to dimension k and $d_{ijk}^{(2)} = 0$.

The FIL model can also be projected into other spaces. One can consider market-level variables $\mathbf{m}$, such as demographics, with, e.g., $\mu_{ij} = \mu\left(\mathbf{d}_{ij}^{(2)}, \mathbf{m}; \gamma, \gamma_m\right)$ or interact them with products characteristics. The projection into price space with $\mu_{ij} = \mu\left(\mathbf{d}_{ij}^{(2)}, \mathbf{d}_{ij}^{\mathbf{p}}; \gamma, \gamma_p\right)$, where $\mathbf{d}_{ij}^{\mathbf{p}}$ is a measure of price distance between products i and j, is left for future research, since it non-trivially changes the shape of the Slutsky matrix, which is not ensured to be symmetric anymore.

4 Empirical Strategy

This section describes the methods to estimate the FIL model with data on market shares s_{jt}, prices p_{jt} and product characteristics $\mathbf{x}_{jt}$ for T markets, indexed by t, and J products per market (see e.g., Berry et al., 1995; Nevo, 2001).

estimated in a nonparametric setting, and on Gandhi and Houde (2017) who use anonymity and symmetry of the ARUM choice probabilities to construct new approximations of the optimal instruments.

[26] This measure is used in the Monte-Carlo simulations of Section 5 and in the empirical application of Section E.

4.1 Estimation by Linear Regression

Nested Logit-Type Linear Regression Using Equations (10) to (13) yields, for each market $t = 1, \ldots, T$,

$$\mu_j \ln\left(s_{jt}\right) + \sum_{i \neq j} \mu_{ij} \ln\left(s_{it} + s_{jt}\right) = \delta_{jt} - c_t, \quad j \in \mathcal{J}_0, \tag{17}$$

and

$$\ln\left(s_{0t}\right) = -c_t, \tag{18}$$

where δ_j is given by Equation (2), and where it is assumed that $\delta_{0t} = 0$ for all markets $t = 1, \ldots, T$.

Combining Equations (17) and (18) with Assumption (i), the FIL model yields the log-odd ratio formula,

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}^{(1)} \beta - \alpha p_{jt} + \sum_{i \neq j} \mu_{ij} \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right) + \xi_{jt}$$

By comparison, for the logit model (8), the log-odd ratio formula is given by

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}^{(1)} \beta - \alpha p_{jt} + \xi_{jt},$$

and, for the three-level nested logit model (9), it is given by

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}^{(1)} \beta - \alpha p_{jt} + \mu_1 \ln\left(\frac{s_{jt}}{\sum_{k \in g} s_{kt}}\right) + \mu_2 \ln\left(\frac{s_{jt}}{\sum_{k \in g|h} s_{kt}}\right) + \xi_{jt},$$

for product j in nest g and subnest $g|h$ (Verboven, 1996).

Then, the FIL model boils down to the nested logit-like linear regression of market shares on product characteristics, prices and shares terms related to the flexible nesting structure.

Price and market shares are endogenous, implying that one needs at least $1 + J(J-1)/2$ instruments, one instrument for price and for each share term. The number of parameters μ_{ij}, equal to $J(J-1)/2$, and thus the number of instruments needed, quickly grows with the number of products.[27] However, the projection of the nesting parameters μ_{ij} into product characteristics space reduces

[27] For $J = 3, 4, 5, 10, 25, 50, 100$, there are $J(J-1)/2 = 3, 6, 10, 45, 300, 1225, 4950$ parameters to be estimated.

the number of parameters, and thus the number of instruments needed.

Projection into Product Characteristics Space The FIL model, projected into product characteristics space, boils down to the following linear IV regression

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}^{(1)}\beta - \alpha p_{jt} + \sum_{i \neq j} \mu\left(\mathbf{d}_{ijt}^{(2)}; \boldsymbol{\gamma}\right) \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right) + \xi_{jt}$$

where the function μ must be specified up to some parameters $\boldsymbol{\gamma}$ to be estimated.

Specifically, I specify μ as a Bernstein polynomial of order M, which, with K_2 product characteristics $\mathbf{x}^{(2)}$, is defined by

$$\mu\left(\mathbf{d}_{ijt}^{(2)}; \boldsymbol{\gamma}\right) = \sum_{k_1=0}^{M} \cdots \sum_{k_{K_2}=0}^{M} \gamma_{\mathbf{k}} B_{\mathbf{k},M}\left(\mathbf{d}_{ijt}^{(2)}\right),$$

where $\mathbf{k} \equiv (k_1, \ldots, k_{K_2})$ and

$$B_{\mathbf{k},M}\left(\mathbf{d}_{ijt}^{(2)}\right) \equiv b_{k_1,M}\left(d_{ijt1}^{(2)}\right) \times \ldots \times b_{k_{K_2},M}\left(d_{ijtK_2}^{(2)}\right)$$

where $b_{k,M}$ is a Bernstein basis function, as defined in Appendix A.3 and where $\boldsymbol{\gamma}$ is the vector of $(M+1)^{K_2}$ parameters. Bernstein polynomials are useful to approximate a continuous function; besides, shape restrictions on μ, such as positivity, monotonicity and concavity, can be easily enforced through linear constraints on parameters $\boldsymbol{\gamma}$ (see e.g. Chak et al., 2005).

Overall, the FIL model projected into product characteristics space is estimated by the following linear IV regression

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}^{(1)}\beta - \alpha p_{jt} + \sum_{\mathbf{k}} \gamma_{\mathbf{k}} \left[\sum_{i \neq j} B_{\mathbf{k},M}\left(\mathbf{d}_{ijt}^{(2)}\right) \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right)\right] + \xi_{jt}. \tag{19}$$

4.2 Optimal Instruments

Assume the existence of exogenous variables $\mathbf{z}_t$ such that the following conditional moment restrictions hold

$$\mathbb{E}\left[\xi_{jt}(\boldsymbol{\theta}) | \mathbf{z}_t\right] = 0, \quad j \in \mathcal{J}_0, \quad t = 1, \ldots, T,$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and where $\xi_{jt} = \xi_{jt}(\boldsymbol{\theta})$ is the jt-product/market unobserved characteristics term defined by Equation (19).

They lead to the following unconditional moment restrictions

$$\mathbb{E}\left[h_{jt}(\mathbf{z}_t)\,\xi_{jt}(\boldsymbol{\theta})\right] = 0, \quad j \in \mathcal{J}_0, \quad t = 1, \ldots, T,$$

where $h_{jt}(\mathbf{z}_t)$ are instruments.

In principle, any function h_{jt} of the exogenous variables $\mathbf{z}_t$ is a candidate as instruments. However, an important practical issue is to choose the ("right") form of h_{jt}, especially to avoid the weak instrument problem.[28]

One solution is to use the optimal instruments of Chamberlain (1987), who shows that the optimal instruments matrix for a single-equation GMM estimator exploit the functional forms of the model and are given by

$$h_{jt}(\mathbf{z}_t) = \mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \boldsymbol{\theta}} | \mathbf{z}_t\right]^{\mathsf{T}} \Omega_{jt}^{-1},$$

where $\Omega_{jt} = \mathbb{E}\left[\xi_{jt}^2 | \mathbf{z}_t\right]$. Following Newey (1990), I consider the homoscedastic case where $\Omega_{jt} = \Omega$, which, in the single-equation GMM estimator case, is a constant that can be set equal to 1 at no loss of generality.

For the FIL model, optimal instruments are given by

$$\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \boldsymbol{\theta}} | \mathbf{z}_t\right]^{\mathsf{T}} = \left(\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \boldsymbol{\beta}^{\mathsf{T}}} | \mathbf{z}_t\right] \quad \mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \alpha} | \mathbf{z}_t\right] \quad \mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \boldsymbol{\gamma}^{\mathsf{T}}} | \mathbf{z}_t\right]\right)^{\mathsf{T}}, \tag{20}$$

where

$$\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \boldsymbol{\beta}^{\mathsf{T}}} | \mathbf{z}_t\right] = -\mathbb{E}\left[\mathbf{x}_t^{(1)} | \mathbf{z}_t\right] = -\mathbf{x}_t^{(1)}, \tag{21}$$

$$\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \alpha} | \mathbf{z}_t\right] = \mathbb{E}\left[p_{jt} | \mathbf{z}_t\right], \tag{22}$$

and

$$\mathbb{E}\left[\frac{\partial \xi_{jt}}{\partial \gamma_{\mathbf{k}}} | \mathbf{z}_t\right] = -\mathbb{E}\left[\sum_{i \neq j} B_{\mathbf{k},M}\left(\mathbf{d}_{ijt}^{(2)}\right) \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right) | \mathbf{z}_t\right], \quad \mathbf{k} \in \{0, \ldots, M\}^{K_2}. \tag{23}$$

[28] Gandhi and Houde (2017) shows in the context of ARUM, that the form of h_j may affect their strength and that choosing the wrong form can lead to the weak instrument problem. Reiss (2016) shows, for the linear regression case, that estimates may be very sensitive to the form of the instruments.

The optimal instruments given by Equations (20) to (23) can be computed in two steps. The first step uses the instruments suggested by Gandhi and Houde (2017) and Pinkse et al. (2002). The second step updates to the optimal instruments of Chamberlain (1987). See Subsection 5.3 for an algorithm that approximates the optimal instruments of the FIL model, based on existing algorithms (Berry et al., 1999; Reynaert and Verboven, 2014; Conlon and Gortmaker, 2018).

5 Performances of the FIL Model

This section studies the ability of the FIL model to provide rich substitution patterns. The RCL model is used for its ability to yield rich substitution patterns (see McFadden and Train, 2000). The RCL and the FIL models are nonnested. Monte Carlo simulations are therefore used to assess its capacity to replicate the substitution patterns of the RCL model (i.e., its own- and cross-price elasticities) for specifications widely used in the literature. To do so, I simulate data from a RCL model and estimate a FIL model to compare the estimated FIL price elasticities of demand to the true RCL price elasticities.[29] [30] Simulations are largely based on Armstrong (2016) and allow to obtain wide ranges of RCL price elasticities.

5.1 Models

Two models are considered. The first model is a structural model of demand with exogenous prices; abstracting from the issue of price endogeneity allows to focus on the performances of the FIL model. The second model is a fully structural model of demand and supply with endogenous prices, i.e., it solves for equilibrium prices and market shares.

The demand side is a standard static RCL model with a single normally distributed random coefficient on an exogenous characteristic.[31] [32] The supply side

[29] This exercise contrasts with Wojcik (2000), who compares the performances of the nested logit and RCL models for out-of-sample prediction of market shares. Instead, I follow Berry and Pakes (2001)'s suggestions by comparing price elasticities.

[30] Another approach would be to compare the estimated FIL price elasticities to the estimated RCL price elasticities. This would recognize that, even in case of no misspecification, the model would be estimated with error (e.g., due to simulation error, strength/validity of the instruments, etc.).

[31] In particular, this specification does not allow for a random coefficient on price.

[32] A future version of this paper will consider the case of several independent and correlated random coefficients. In particular, it could be interesting to investigate how the FIL model performs with respect to the RCL with independent random coefficients when the random coeffi-

is a static oligopolistic price competition model with multiproduct firms.

Demand The conditional indirect utility of consumer n in market t from choosing an inside product j is given by

$$u_{njt} = \beta_0 + \beta_{nx} x_{jt} - \alpha p_{jt} + \xi_{jt} + \varepsilon_{njt}, \tag{24}$$

where x_{jt} is the exogenous characteristic, p_{jt} is the price, and ξ_{jt} is the unobserved product characteristic of product j in market t. The utility from choosing the outside good $j = 0$ is normalized to $u_{n0t} = \varepsilon_{n0t}$, for all markets $t = 1, \ldots, T$.

The two parameters β_0 and α are assumed to be equal for all consumers n: β_0 captures the value of choosing an inside product instead of the outside good, and α is the marginal utility of income. The parameter β_{nx} is the only random coefficient, which captures consumer-specific valuation for characteristic x_{jt}. The term ε_{njt} is a remaining consumer-specific valuation for product j. In the RCL model, the ε_{njt}'s are assumed to be distributed i.i.d. type I extreme value.

The random coefficient can then be decomposed as $\beta_{nx} = \beta_x + \sigma_x v_{nx}$, where v_{nx} is a standardized random variable (i.e., with mean 0 and variance 1), so that β_x captures the mean valuation for product characteristic x_{jt} and σ_x is its standard deviation across consumers. Then, the indirect utility (24) can be rewritten as

$$u_{njt} = \beta_0 + \beta_x x_{jt} - \alpha p_{jt} + \xi_{jt} + \sigma_x x_{jt} v_{nx} + \varepsilon_{njt}.$$

Each consumer n with random preferences β_{nx} chooses one unit of the product that provides her the highest utility. Then, the market share of product j in market t is computed as the probability that product j provides the highest utility across all products in market t. In the RCL model, it is given by

$$s_{jt} = \sigma_j(\mathbf{x}_t, \mathbf{p}_t, \boldsymbol{\xi}_t; \boldsymbol{\theta}) = \int \frac{\exp(\beta_0 + \beta_x x_{jt} - \alpha p_{jt} + \xi_{jt} + \sigma_x x_{jt} v)}{1 + \sum_{k=1}^{J} \exp(\beta_0 + \beta_x x_{kt} - \alpha p_{kt} + \xi_{kt} + \sigma_x x_{kt} v)} f(v) \mathrm{d}v, \tag{25}$$

and

$$s_{0t} = \sigma_0(\mathbf{x}_t, \mathbf{p}_t, \boldsymbol{\xi}_t; \boldsymbol{\theta}) \equiv 1 - \sum_{k=1}^{J} \sigma_k(\mathbf{x}_t, \mathbf{p}_t, \boldsymbol{\xi}_t; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} \equiv (\beta_0, \alpha, \beta_x, \sigma_x)$ and where $f(v)$ denotes the distribution of v_{nx}.

cients are actually correlated.

Supply Consider F firms, indexed by f. The profit that firm f producing the set of products $\mathcal{J}_f$, with $\cup_{f=1}^{F}\mathcal{J}_f = \mathcal{J}_0$ and $\cap_{f=1}^{F}\mathcal{J}_f = \emptyset$, makes in market t is given by

$$\Pi_{ft} = \sum_{j\in\mathcal{J}_f}\left(p_{jt} - c_{jt}\right)\sigma_j(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \boldsymbol{\theta}), \tag{26}$$

where c_{jt}, the marginal cost, is parametrized as follows

$$c_{jt} = \gamma_0 + \gamma_x x_{jt} + \gamma_w w_{jt} + \omega_{jt},$$

where x_{jt} is the product characteristic, which affects utility and cost, w_{jt} is a variable which only affects cost, and ω_{jt} is an unobserved cost component.

In the static oligopolistic price competition model, each firm f in market t chooses the prices p_{jt} of its products $j \in \mathcal{J}_f$ to maximize its profits (26), given the characteristics and costs of its products and the prices, characteristics and costs of its competing products in that market. Assuming that a pure-strategy Nash equilibrium exists, prices $\mathbf{p}_t \equiv (p_{1t}, \ldots, p_{Jt})$ solve the following first-order conditions

$$\boldsymbol{\sigma}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \boldsymbol{\theta}) + \nabla_{\mathbf{p}_t}(\mathbf{p}_t - \mathbf{c}_t) = \mathbf{0}_J, \quad t = 1, \ldots, T, \tag{27}$$

where $\nabla_{\mathbf{p}_t}$ has entries ij given by

$$-\frac{\partial\sigma_i}{\partial p_{jt}}(\mathbf{x}_t, \mathbf{p}_t, \xi_t; \boldsymbol{\theta})\,\Theta_{ij}, \tag{28}$$

with $\Theta_{ij} = 1$ if products i and j are produced by the same firm, and $\Theta_{ij} = 0$ otherwise. Note that this model rules out price coordination among firms (see e.g., Michel and Weiergraeber, 2019, for a model allowing for price coordination).

5.2 Simulation Configurations

The simulations are largely based on those of Armstrong (2016). For each configuration, I construct 500 Monte Carlo datasets, and for each of them, I simulate T markets, where each one consists of J products, N consumers (indexed by n), and $J/10$ firms (indexed by f) producing each one 10 products.

Each product j in market t is characterized by the vector $(s_{jt}, p_{jt}, x_{jt}, \xi_{jt}, w_{jt}, \omega_{jt})$, where x_{jt} and w_{jt} are drawn from two independent standard uniform distributions, and where the structural error terms, ξ_{jt} and ω_{jt}, are computed as

$$\xi_{jt} = u_{jt}^1 + u_{jt}^3 - 1,$$
$$\omega_{jt} = u_{jt}^1 + u_{jt}^2 - 1,$$

where u_{jt}^1, u_{jt}^2 and u_{jt}^3 are drawn from three independent standard uniform distributions.

Exogenous Prices In the model with exogenous prices, prices are computed as

$$p_{jt} = 0.5\eta_{jt} + 4,$$

where η_{jt} is drawn from a standard normal distribution, so that the distribution of prices thus obtained is very similar to that obtained in the model with endogenous prices. Market shares (25) are computed as follows. I first generate $N = 2,000$ draws v_{nx} from a standard normal distribution. Then, I compute the observed market share of product j in market t as

$$s_{jt}(\mathbf{x}_t, \mathbf{p}_t, \boldsymbol{\xi}_t | \boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N} \frac{\exp(\beta_0 + \beta_x x_{jt} - \alpha p_{jt} + \xi_{jt} + \sigma_x x_{jt} v_{nx})}{1 + \sum_{k=1}^{J} \exp(\beta_0 + \beta_x x_{kt} - \alpha p_{kt} + \xi_{kt} + \sigma_x x_{kt} v_{nx})}.$$

I consider 16 configurations by varying $\sigma_x \in \{0.5, 1, 2, 3\}$, $J \in \{25, 50\}$ and $T \in \{25, 50\}$, with $(\alpha, \beta_0, \beta_x, \gamma_0, \gamma_x, \gamma_w)$ fixed at $(1, 3, 6, 2, 1, 1)$.

Endogenous Prices In the model with endogenous prices, for each market t, prices $\mathbf{p}_t$ and market shares $\mathbf{s}_t \equiv \left(s_{1t}, \ldots, s_{Jt}\right)$ solve the structural model of demand and supply, i.e., solve Equations (25) and (27). I consider 4 configurations by varying $\sigma_x \in \{0.5, 1, 2, 3\}$, with $(\alpha, \beta_0, \beta_x, \gamma_0, \gamma_x, \gamma_w)$ fixed at $(1, 3, 6, 2, 1, 1)$.

5.3 Optimal Instruments

This subsection describes the algorithms used to compute the optimal instruments of the FIL model given by Equations (20) to (23).

Exogenous Prices The first step uses instruments suggested by Pinkse et al. (2002) and Gandhi and Houde (2017). The included instruments are the constant, x_j and p_j. The excluded instruments are of the form of

$$\sum_{i \neq j} b_{k,M}\left(d_{ijt}\right) z_{ijt}, \quad k = 0, \ldots, M, \tag{29}$$

where $d_{ijt} \equiv |x_{jt} - x_{it}|$ and $z_{ijt} \in \{x_{jt}, p_{jt}, x_{jt} - x_{it}, p_{jt} - p_{it}\}$ explain much of the variation in $\ln\left(s_{jt}/(s_{it} + s_{jt})\right)$, i.e., the relative popularity of product j with respect to product i.

The second step updates to (an approximation of) the optimal instruments of Chamberlain (1987), which are given by

$$\sum_{i \neq j} b_{k,M}\left(d_{ijt}\right) \ln\left(\frac{\widehat{s}_{jt}}{\widehat{s}_{it} + \widehat{s}_{jt}}\right), \quad k = 0, \ldots, M, \tag{30}$$

where $\widehat{s}_{jt}$ is the predicted market shares of product j in market t. The prediction uses the estimated parameters of the first step and is evaluated at the expected value of the unobservables, $\mathbb{E}[\xi_{jt}] = 0$.

Endogenous Prices For the model with endogeneous prices, the algorithm is a little bit different and exploits the supply side (in particular, the ownership structure of the market).

The first step estimates the demand model and the marginal cost function. For demand estimation, the included instruments are the constant and x_{jt}, and the excluded instruments are of the form of Equation (29), where $z_{ijt} \in \{x_{jt}, x_{jt} - x_{it}\}$, and $\widehat{p}_{jt}$ computed as the predicted value of the linear regression of p_{jt} on x_j, w_j, x_j^2, w_j^2, $x_j w_j$, $\sum_{k \in \mathcal{J}_f(j)\setminus\{j\}} x_{jt}$, $\sum_{k \in \mathcal{J}\setminus\mathcal{J}_f(j)} x_{jt}$, $\sum_{k \in \mathcal{J}_f(j)\setminus\{j\}} w_{jt}$, $\sum_{k \in \mathcal{J}\setminus\mathcal{J}_f(j)} w_{jt}$. Estimation leads to demand estimates $\widehat{\boldsymbol{\theta}} \equiv (\widehat{\alpha}, \widehat{\beta}_0, \widehat{\beta}_x, \widehat{\gamma})$, from which one can compute $\widehat{v}_{jt} = \widehat{\beta}_0 + \widehat{\beta}_x x_{jt}$.

The implied marginal cost c_{jt} are obtained from the first-order conditions (27) and (28) where $\boldsymbol{\sigma}$ defined by its inverse given by Equations (10) and (11). The predicted marginal costs are obtained as the predicted value of the linear regression of c_{jt} on a constant, x_{jt} and w_{jt}: $\widehat{c}_{jt} = \widehat{\gamma}_0 + \widehat{\gamma}_x x_{jt} + \widehat{\gamma}_w w_{jt}$.

The second step uses $\widehat{p}_{jt}$ and instruments given by Equation (30) as optimal instruments, where $\widehat{p}_{jt}$ and $\widehat{s}_{jt}$ solve the first-order conditions as a function of $\mathbf{x}$, $\widehat{\mathbf{v}}$, $\widehat{\mathbf{c}}$, and $\widehat{\boldsymbol{\theta}}$. The prediction uses the estimated parameters of the first step and is evaluated at the expected value of the unobservables, $\mathbb{E}[\xi_{jt}] = \mathbb{E}[\omega_{jt}] = 0$.

5.4 Results

This subsection studies the ability of the FIL model to replicate the substitution patterns the RCL model for the configurations described in the previous sections.

For each configuration, μ is specified as a Bernstein polynomial of order 5. Results are summarized in Tables 1 and 2 for the 16 configurations with exogenous prices and in Table 3 for the 4 configurations with endogenous prices. Elasticities are computed based on the parameters estimates as given by the mean estimated value of parameters across Monte-Carlo datasets, keeping all market-level variables at the values of the first Monte-Carlo dataset and using the first twenty markets (for the sake of comparability between the configurations).

Consider first the results for the configurations with exogenous prices. They confirm the ability of the FIL model to match the own- and cross-price from the RCL models. The lower σ_x is, the better the fit is. Models with a higher number of products ($J = 50$ versus $J = 25$) are better fitted, while the number of market T does not significantly affect the results. The average value of the true own-price elasticities ranges from -3.9021 to -3.8015 and the FIL model estimated these elasticities with a bias ranging from -0.0132 to -0.2219. The average value of the true cross-price elasticities ranges 0.0736 to 0.1520 and the FIL model estimated these elasticities with a bias ranging from 0.0007 to 0.0139.

Turn now to the configurations with endogeneity. The FIL model provides a good approximation – except maybe for the configuration with $\sigma_x = 3$ where the MSE is high. The average value of the true own-price elasticities ranges from -4.1208 to -4.0894 and the FIL model estimated these elasticities with a bias ranging from 0.0207 to -0.1065. The average value of the true cross-price elasticities ranges 0.1721 to 0.1855 and the FIL model estimated these elasticities with a bias ranging from 0.0019 to 0.0109.

Several remarks are in order. First, the FIL model does not generate complements as defined by a negative cross-price derivative of demand, when, as it is the case of the RCL model, there is only substitutes. then, Appendix D provides further results about the simulations. Figures 1 to 5 represent the accuracy of the estimation of the FIL model when the true model is the RCL model (after removing outliers). They show the ability of the FIL model to mimic the substitution patterns of the simulated RCL models. Tables 4 to 6 summarize the results obtained using cruder instruments. They show the usefulness of the optimal instruments for fitting, since these latter allow to obtain better fits.

Table 1: Simulation Results: Own-Price Elasticities with Exogenous Prices

T	J	σ	Mean		Percentiles				Bias	MSE
					2.5th	97.5th	2.5th	97.5th		
			RCL	FIL	RCL		FIL			
20	25	0.5	-3.8338	-3.8769	-4.8205	-2.5582	-5.0968	-2.3005	-0.0430	0.0323
50	25	0.5	-3.8333	-3.8695	-4.8196	-2.5142	-5.0571	-2.3286	-0.0362	0.0341
20	50	0.5	-3.9021	-3.9154	-4.9011	-2.8028	-5.0224	-2.6916	-0.0132	0.0169
50	50	0.5	-3.9021	-3.9185	-4.8997	-2.7943	-5.0409	-2.6710	-0.0163	0.0190
20	25	1	-3.8314	-3.8863	-4.8191	-2.5578	-5.1908	-2.2778	-0.0549	0.0445
50	25	1	-3.8309	-3.8795	-4.8179	-2.5121	-5.1498	-2.3064	-0.0485	0.0477
20	50	1	-3.9009	-3.9183	-4.9014	-2.7982	-5.1053	-2.6696	-0.0174	0.0225
50	50	1	-3.9009	-3.9217	-4.9000	-2.7932	-5.0999	-2.6460	-0.0208	0.0254
20	25	2	-3.8201	-3.9228	-4.8144	-2.5156	-5.5502	-2.2665	-0.1027	0.1201
50	25	2	-3.8196	-3.9194	-4.8122	-2.5135	-5.5527	-2.2320	-0.0998	0.1362
20	50	2	-3.8948	-3.9310	-4.9015	-2.7790	-5.3540	-2.6058	-0.0362	0.0660
50	50	2	-3.8947	-3.9367	-4.8997	-2.7719	-5.3891	-2.5588	-0.0420	0.0768
20	25	3	-3.8018	-4.0176	-4.7905	-2.4682	-6.6121	-2.1862	-0.2158	0.4683
50	25	3	-3.8015	-4.0234	-4.8003	-2.5089	-6.6773	-2.1090	-0.2219	0.5479
20	50	3	-3.8839	-3.9761	-4.8957	-2.7562	-6.0543	-2.4848	-0.0922	0.2694
50	50	3	-3.8837	-3.9852	-4.8922	-2.7618	-6.1047	-2.4542	-0.1015	0.3069

Notes: Summary statistics across 500 Monte Carlo replications. The bias is measured by the mean error. The mean square error (MSE) measures the accuracy.

Table 2: Simulation Results: Cross-Price Elasticities with Exogenous Prices

T	J	σ	Mean		Percentiles				Bias	MSE
					2.5th	97.5th	2.5th	97.5th		
			RCL	FIL	RCL		FIL			
20	25	0.5	0.1515	0.1542	0.0011	0.9628	0.0006	0.9399	0.0027	0.0004
50	25	0.5	0.1520	0.1544	0.0011	0.9060	0.0007	0.9208	0.0025	0.0004
20	50	0.5	0.0752	0.0759	0.0004	0.4714	0.0003	0.4866	0.0007	0.0001
50	50	0.5	0.0752	0.0760	0.0005	0.5038	0.0003	0.5023	0.0007	0.0001
20	25	1	0.1512	0.1546	0.0016	0.9224	0.0008	0.9153	0.0034	0.0005
50	25	1	0.1517	0.1549	0.0016	0.8845	0.0008	0.8875	0.0032	0.0005
20	50	1	0.0752	0.0760	0.0007	0.4548	0.0003	0.4710	0.0008	0.0001
50	50	1	0.0752	0.0761	0.0007	0.4820	0.0003	0.4902	0.0009	0.0001
20	25	2	0.1496	0.1560	0.0037	0.8192	0.0014	0.8135	0.0063	0.0012
50	25	2	0.1501	0.1564	0.0038	0.7880	0.0015	0.7668	0.0063	0.0013
20	50	2	0.0748	0.0763	0.0017	0.4083	0.0006	0.4235	0.0015	0.0004
50	50	2	0.0748	0.0764	0.0017	0.4216	0.0007	0.4468	0.0016	0.0004
20	25	3	0.1459	0.1594	0.0062	0.7109	0.0026	0.6754	0.0135	0.0034
50	25	3	0.1463	0.1601	0.0062	0.6927	0.0026	0.6327	0.0139	0.0038
20	50	3	0.0736	0.0771	0.0029	0.3514	0.0011	0.3684	0.0035	0.0009
50	50	3	0.0736	0.0773	0.0030	0.3696	0.0012	0.3886	0.0037	0.0010

Notes: Summary statistics across 500 Monte Carlo replications. The bias is measured by the mean error. The mean square error (MSE) measures the accuracy.

Table 3: SIMULATION RESULTS WITH ENDOGENOUS PRICES

T	J	σ	Mean		Percentiles				Bias	MSE
					2.5th	97.5th	2.5th	97.5th		
			RCL	FIL	RCL		FIL			
Own-Price Elasticities										
50	25	0.5	-4.1208	-4.1001	-5.4317	-2.6397	-5.4778	-2.6550	0.0207	0.0513
50	25	1	-4.1178	-4.0973	-5.4325	-2.6380	-5.6391	-2.6395	0.0205	0.0938
50	25	2	-4.1060	-4.0612	-5.4303	-2.6084	-6.4735	-2.4461	0.0448	0.3925
50	25	3	-4.0887	-4.1952	-5.4167	-2.5384	-9.2503	-2.2480	-0.1065	1.7778
Cross-Price Elasticities										
50	25	0.5	0.1855	0.1874	0.0038	0.8959	0.0022	0.9031	0.0019	0.0007
50	25	1	0.1843	0.1862	0.0050	0.8627	0.0026	0.8472	0.0020	0.0008
50	25	2	0.1792	0.1812	0.0087	0.7678	0.0037	0.6914	0.0020	0.0027
50	25	3	0.1720	0.1829	0.0110	0.7035	0.0042	0.5806	0.0109	0.0087

Notes: Summary statistics across 500 Monte Carlo replications. The bias is measured by the mean error. The mean square error (MSE) measures the accuracy.

6 Conclusion

This paper has developed the FIL model, which is an inverse demand models for differentiated products that accommodates rich substitution patterns, including complementarity, thanks to a simple linear regression with data on market shares, prices and product characteristics.

The FIL model uses a flexible nesting structure with a nest for each pair of products. By contrast with the nested logit model, nesting in the FIL model is just a way to fully parametrize the matrix of price elasticities of demand and does not require the modeller to choose the nesting structure before estimation. This means that the resulting cross-price elasticities of demand are not constrained by the model, but instead are driven by the data.

The FIL model, projected into product characteristics space, makes the price elasticities depending on product characteristics directly, as it is the case of the "flexible" RCL model. On top of that, as shown using Monte Carlo simulations, it is able to mimic the substitution patterns from the RCL model.

The FIL model can be applied to various topics in industrial organization, international trade, public economics, etc. It can be used to answer relevant policy questions, such as the effect of mergers, products' entry, and changes in regulation. Due to its simplicity of estimation, the likely audience of the FIL model involves researchers as well as antitrust practitioners in consultancies and competition authorities who wish to avoid complex procedures of estimation and/or who are under time pressure.

Appendices

A Preliminaries

This section summarizes useful results and definitions.

A.1 Demand Invertibility

Proposition 4. Assume that *(i)* $\mathbf{G} = \left(G_0, \ldots, G_J\right) : [0, \infty)^{J+1} \to [0, \infty)^{J+1}$ is continuously differentiable and homogeneous of degree one on $\text{int}\left(\Delta_J\right)$; *(ii)* $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}(\mathbf{s})$ is positive definite and symmetric on $\text{int}\left(\Delta_J\right)$; and *(iii)* $|\ln \mathbf{G}(\mathbf{s})|$ approaches infinity as $\mathbf{s}$ approaches $\text{bd}(\Delta_J)$.

(i) It follows that $\ln \mathbf{G}$ is invertible on $\text{int}(\Delta_J)$.

(ii) Consider any vector of market shares $\mathbf{s} \in \text{int}\left(\Delta_J\right)$. Then, there exists a unique $\delta \in \mathbb{R}_0^{J+1}$ such that

$$\delta = \sigma^{-1}(\mathbf{s}) \Leftrightarrow \mathbf{s} = \sigma(\delta).$$

Proof. See Proposition 2 in Monardo (2019). □

A.2 Flexibility of Demands

Consider a demand system of $(J+1)$ differentiated products $\sigma(\mathbf{p})$. Absent income effect, this demand system is said to be flexible in the sense of Diewert (1974) if it is able to provide a first-order approximation to any theoretically grounded demand system at a point in price space, i.e., if it can match the $(J+1)$ true (i.e., observed) market shares $\mathbf{s}^*$, and the $(J+1)^2$ true own- and cross-price elasticities.[33]

[33] See Barnett (1983, 1985) for the different definitions of flexibility and their relationship to second-order approximations.

However, only $J + J(J+1)/2$ independent market shares and elasticities need to be matched. Indeed, since the market shares sum to one, only J market shares are independent. Once the market shares of J products are matched, the market shares of the $(J+1)$th product is matched automatically using that market shares sum to one. Then, given the market shares, by Slutsky symmetry, only $J(J+1)/2$ cross-price elasticities are independent. Lastly, given the market shares and the cross-price elasticities, one can infer all the own-price elasticities from

$$\frac{\partial \sigma_i(\mathbf{p})}{\partial p_i} = -\sum_{j \neq i} \frac{\partial \sigma_j(\mathbf{p})}{\partial p_i}, \quad i \in \mathcal{J},$$

which is obtained by differentiating $\sum_{j \in \mathcal{J}} \sigma_j(\mathbf{p}) = 1$ with respect to price p_i.

A.3 Bernstein Polynomials

For a positive integer M, the Bernstein basis functions defined over interval $[a,b]$ are defined by

$$b_{k,M}(x) \equiv \binom{M}{k} \frac{(x-a)^k (b-x)^{M-k}}{(b-a)^M},$$

where $k = 0, \ldots, M$.[34] In applications, Bernstein basis functions are often expressed over interval $[0,1]$ as

$$b_{k,M}(x) = \binom{M}{k} x^k (1-x)^{M-k}.$$

A univariate function defined over interval $[a,b]$ can be approximated by a linear combination of the Bernstein basis functions

$$\sum_{k=0}^{M} \theta_k b_{k,M}(x),$$

for $x \in [a,b]$ and for some coefficients θ_k, $k = 0, \ldots, M$.

The generalization to a multivariate function defined over $[a_1, b_1] \times \ldots \times [a_L, b_L]$

[34] See Chapter 6 in Davis (1975) for more on Bernstein polynomials.

is straightforward. A multivariate function can be approximated by

$$\sum_{k_1=0}^{M} \cdots \sum_{k_L=0}^{M} \theta_{k_1,\ldots,k_L} b_{k_1,M}(x_1) \times \ldots \times b_{k_L,M}(x_L),$$

for $(x_1,\ldots,x_L) \in [a_1,b_1] \times \ldots \times [a_L,b_L]$ and for some coefficients $\theta_{k_1,\ldots,k_L}$, $k_1 = 0,\ldots,M$, $\ldots$, $k_L = 0,\ldots,M$.

The derivative of $b_{k,M}(x)$ is given by

$$\frac{\partial b_{k,M}(x)}{\partial x} = M\left[b_{k-1,M-1}(x) - b_{k,M-1}(x)\right],$$

that is,

$$\frac{\partial b_{k,M}(x)}{\partial x} = \left[\frac{k}{x-a} - \frac{M-k}{b-x}\right] b_{k,M}(x),$$

which implies that approximation also works for the derivatives (see Chak et al., 2005).

A.4 Defining Complementarity and Substitutability

There are different ways of defining complementarity (Samuelson, 1974; Newman, 2008). The empirical literature has focused on two different but related definitions, both relying on the idea of a positive interaction between the products.

The first definition asserts that products i and j are complements (resp., substitutes) in utility if the cross second-order partial derivative of the utility function with respect to quantities s_i and s_j

$$\frac{\partial^2 u(\mathbf{s})}{\partial s_i \partial s_j} = \frac{\partial^2 u(\mathbf{s})}{\partial s_j \partial s_i}$$

is positive (resp., negative).

Absent income effect, the second definition, which is the textbook definition, asserts that products i and j are complements (resp., substitutes) in demand if the cross-price derivative of demand[35]

$$\frac{\partial \sigma_i(\delta)}{\partial p_j} = \frac{\partial \sigma_j(\delta)}{\partial p_i}$$

is negative (resp., positive).[36]

This appendix does not discuss the advantages/disadvantages of both definitions. However, just note that the utility-based definition has been criticized on the ground that it is not invariant with respect to monotone increasing transformations of utility u. Regarding the demand-based definition, whether two products are complements or substitutes depends on the relation of the two products to the other products, which may lead to a wrong classification (Samuelson, 1974; Ogaki, 1990).

B Proofs

B.1 Proof of Proposition 1

The FIL model is defined by $\ln \mathbf{G} \equiv (\ln G_0, \ldots, \ln G_J)$ with

$$G_j(\mathbf{s}; \boldsymbol{\mu}) \equiv s_j^{\mu_j} \prod_{i \neq j} \left(s_i + s_j\right)^{\mu_{ij}}, \quad j \in \mathcal{J}_0, \tag{31}$$

and

$$G_0(\mathbf{s}) = s_0, \tag{32}$$

where $\boldsymbol{\mu} = \left(\mu_1, \ldots, \mu_J, \mu_{11}, \ldots, \mu_{J-1,J}\right)$ satisfies Assumptions (i) – (iii).

For the proof, it is convenient to use the conventions $\mu_0 = 1$ and $\mu_{j0} = \mu_{0j} = 0$ for all $j \in \mathcal{J}_0$.

[35]In the standard definition of complementarity, two products are complements (resp., substitutes) in demand if the compensative cross-price derivative of demand is negative (resp., positive). Absent income effect, the Slutsky matrix is just the matrix of own- and cross-price derivatives of demand, and in turn, complementarity (resp., substitutability) is the only source of the negative (resp., positive) cross-price derivative.

[36]The utility-based definition is known as the Edgeworth-Pareto (EP) or Auspitz-Lieben-Edgeworth-Pareto (ALEP) definition of complementarity; the demand-based definition is known as the Slutsky-Hicks-Allen-Schultz (SHAS) definition of complementarity.

The Jacobian $\mathbf{J}_{\ln \mathbf{G}}$ has diagonal entries $j+1, j+1$ given by

$$\frac{\mu_j}{s_j} + \sum_{i \neq j} \frac{\mu_{ij}}{s_i + s_j},$$

and off-diagonal entries $i+1, j+1$ given by

$$\frac{\mu_{ij}}{s_i + s_j}.$$

The proof uses Proposition 4 as applied to the function $\mathbf{G}(.,\boldsymbol{\mu})$ given by Equations (31) and (32). It thus consists in showing that the function $\mathbf{G}(.,\boldsymbol{\mu})$ is homogeneous of degree one and that the Jacobian $\mathbf{J}_{\ln \mathbf{G}}$ is positive definite and symmetric.

It is homogeneous of degree one, since for $\lambda > 0$ and $j \in \mathcal{J}_0$,

$$\begin{aligned}
G_j(\lambda \mathbf{s}) &= \left(\lambda s_j\right)^{\mu_j} \prod_{i \neq j} \left[\lambda (s_i + s_j)\right]^{\mu_{ij}}, \\
&= \left[\lambda^{\mu_j} \prod_{i \neq j} \lambda^{\mu_{ij}}\right] \left[\left(s_j\right)^{\mu_j} \prod_{i \neq j} \left(s_i + s_j\right)^{\mu_{ij}}\right], \\
&= \left[\lambda^{\mu_j + \sum_{i \neq j} \mu_{ij}}\right] G_j(\mathbf{s}), \\
&= \lambda G_j(\mathbf{s}),
\end{aligned}$$

where the last equality uses Assumption (i). Likewise, for $\lambda > 0$,

$$G_0(\lambda \mathbf{s}) = \lambda s_0 = \lambda G_0(\mathbf{s}).$$

The Jacobian $\mathbf{J}_{\ln \mathbf{G}}$ is symmetric since, by Assumption (iii), its entry $i+1, j+1$ given by

$$\frac{\mu_{ij}}{s_i + s_j}$$

is equal to its entry $j+1, i+1$ given by

$$\frac{\mu_{ji}}{s_j + s_i}.$$

The Jacobian $\mathbf{J}_{\ln \mathbf{G}}$ is positive definite. To show this, let $\langle i, j \rangle$ be a vector with entry k given by

$$\langle i,j\rangle_k = \begin{cases} 1 & k=i, \\ 1 & k=j, \\ 0 & \text{otherwise,} \end{cases}$$

$\langle i,-j\rangle$ be a vector with entry k given by

$$\langle i,-j\rangle_k = \begin{cases} 1 & k=i \\ -1 & k=j \\ 0 & \text{otherwise,} \end{cases}$$

and $\langle i\rangle$ be a vector with entry k given by

$$\langle i\rangle_k = \begin{cases} 1 & k=i, \\ 0 & \text{otherwise.} \end{cases}$$

Then, using Assumption (iii), one can write the Jacobian $\mathbf{J}_{\ln \mathbf{G}}$ as

$$\sum_{i\in\mathcal{J}} \frac{\mu_i \langle i\rangle\langle i\rangle^\intercal}{s_i} + \sum_{0<i<j,\mu_{ij}>0} \frac{|\mu_{ij}|\langle i,j\rangle\langle i,j\rangle^\intercal}{s_i+s_j} + \sum_{0<i<j,\mu_{ij}<0} \frac{|\mu_{ij}|\langle i,-j\rangle\langle i,-j\rangle^\intercal}{s_i+s_j},$$

which is positive definite since its first term is a positive definite matrix by Assumption (ii), and its second and third terms are two sums of positive semi-definite matrices.

B.2 Proof of Inverse Slutsky Matrix

Consider the class of inverse demands given by

$$\sigma_j^{-1}(\mathbf{s}) \equiv \ln G_j(\mathbf{s}) + c = \delta_j, \quad j\in\mathcal{J}_0,$$

and

$$\sigma_0^{-1}(\mathbf{s}) \equiv \ln(s_0) + c = \delta_0,$$

where $c\in\mathbb{R}$ is a market-specific constant, where $\mathbf{G}$ is homogeneous of degree one and where the Jacobian $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ is positive definite and symmetric. This is a subclass of the class of GIL models, developed by Fosgerau et al. (2019).

For this class of inverse models,

$$\left[\mathbf{J}_{\sigma}^{\delta}(\delta)\right]_0 = \left[\mathbf{J}_{\ln\mathbf{G}}(\mathbf{s})\right]_0^{-1}\left[\mathbf{I}_J - \mathbf{1}_J\mathbf{s}_{-0}^{\mathsf{T}}\right];$$

The inverse demand being the target of estimation, it is worthwhile to derive the expression of the matrix of inverse demand derivatives, which, by the implicit function theorem, satisfies

$$\left[\mathbf{J}_{\sigma^{-1}}^{\mathbf{s}}(\mathbf{s})\right]_0 = \left[\mathbf{J}_{\sigma}^{\delta}(\delta)\right]_0^{-1},$$

where $\mathbf{s} = \sigma(\delta)$. Then,

$$\left[\mathbf{J}_{\sigma^{-1}}^{\mathbf{s}}(\mathbf{s})\right]_0 = \left[\mathbf{I}_J - \mathbf{1}_J\mathbf{s}_{-0}^{\mathsf{T}}\right]^{-1}\left[\mathbf{J}_{\ln\mathbf{G}}(\mathbf{s})\right]_0,$$

where

$$\left[\mathbf{I}_J - \mathbf{1}_J\mathbf{s}_{-0}^{\mathsf{T}}\right]^{-1} = \frac{1}{s_0}\begin{bmatrix} s_0+s_1 & s_2 & \cdots & s_J \\ s_1 & s_0+s_2 & \cdots & s_J \\ \vdots & \vdots & \ddots & \vdots \\ s_1 & s_2 & \cdots & s_0+s_J \end{bmatrix}.$$

Noting that $\frac{\partial \ln G_0(\mathbf{s})}{\partial s_j} = \frac{\partial \ln G_j(\mathbf{s})}{\partial s_0} = 0$ for all $j \in \mathcal{J}_0$, one obtains the requested expression (14) using $\sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} = 1$ for all $k \in \mathcal{J}_0$, and that $\left[\mathbf{J}_{\sigma}^{\mathbf{p}}(\delta)\right]_0 = -\alpha\left[\mathbf{J}_{\sigma}^{\delta}(\delta)\right]_0$.

B.3 Proof of Proposition 2

The proof consists in showing that, observing the vector of market shares $\mathbf{s}^*$, the FIL model can match that vector of market shares $\mathbf{s}^*$ as well as the true (symmetric) matrix of own- and cross-price derivatives of demand for inside products $\left[\mathbf{J}_{\mathbf{s}}^{\mathbf{p}^*}\right]_0$. In other words, one must show that one can choose δ^*, α^* and μ^* so that

$$\mathbf{s}^* = \sigma(\delta^*; \mu^*), \tag{33}$$

and

$$\left[\mathbf{J}_{\mathbf{s}}^{\mathbf{p}*}\right]_0 = \left[\mathbf{J}_{\sigma}^{\mathbf{p}}(\delta^*; \mu^*)\right]_0, \tag{34}$$

where σ is defined by its inverse in Equations (10) to (13).

Regarding the market shares, Equation (33) is invertible, with inverse given by

$$\delta_j^* = \sigma_j^{-1}(\mathbf{s}^*;\boldsymbol{\mu}^*) = \ln G_j(\mathbf{s}^*;\boldsymbol{\mu}^*) + c,$$

where $\ln G_j$ is given by Equation (11), and

$$\delta_0^* = \ln(s_0^*) + c.$$

Setting $\delta_0^* = 0$, one obtains $c = -\ln\left(s_0^*\right)$, so that, to match the vector of market shares $\mathbf{s}^*$, one can set

$$\delta_j^* = \ln G_j(\mathbf{s}^*;\boldsymbol{\mu}^*) - \ln(s_0^*),$$

that is

$$\delta_j^* = \ln\left(\frac{s_j^*}{s_0^*}\right) - \sum_{i \neq j} \mu_{ij}^* \ln\left(\frac{s_j^*}{s_i^* + s_j^*}\right).$$

Regarding the price derivatives, Equation (34) given by

$$\left[\mathbf{J}_{\mathbf{s}}^{\mathbf{p}^*}\right]_0 = -\alpha^* [\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s}^*)]_0^{-1} \left(\mathbf{I}_J - \mathbf{1}_J \mathbf{s}_{-0}^{*\top}\right),$$

with $\mathbf{s}^* = \boldsymbol{\sigma}(\boldsymbol{\delta}^*;\boldsymbol{\mu}^*)$, can be inverted to give

$$[\mathbf{J}_{\ln \mathbf{G}}(\mathbf{s}^*;\boldsymbol{\mu}^*)]_0 = -\alpha^* \left(\mathbf{I}_J - \mathbf{1}_J \mathbf{s}^{*\top}\right) \left[\mathbf{J}_{\mathbf{s}_{-0}}^{\mathbf{p}^*}\right]_0^{-1} \equiv \alpha^* \boldsymbol{\Gamma}^*(\mathbf{s}^*),$$

that is, for entry ij,

$$\frac{\partial \ln G_i(\mathbf{s}^*)}{\partial s_j} = \frac{\mu_{ij}^*}{s_i^* + s_j^*} = \alpha^* \Gamma_{ij}^*(\mathbf{s}^*),$$

where Γ_{ij}^* is the entry ij of the matrix $\boldsymbol{\Gamma}^*$. This implies that, to match the off-diagonal elements, one can set

$$\mu_{ij}^* = \alpha^* \left(s_i^* + s_j^*\right) \Gamma_{ij}^*(\mathbf{s}^*).$$

Once the off-diagonal elements are matched, the diagonal elements are automatically, since, the FIL model satisfies the Euler-type equation

$$\sum_{k \in \mathcal{J}} \frac{\partial \ln G_j(\mathbf{s}^*)}{\partial s_k} s_k^* = 1, \quad j \in \mathcal{J}.$$

Lastly, the parameter α^* can be chosen so that $\mu_j > 0$ for all $j \in \mathcal{J}_0$.

C Projection into Product Characteristics Space

Additive Random Utility Models (ARUM) Since McFadden (1974), the ARUM has been employed in many fields of economics (see Table 1 in Berry and Haile, 2016). In particular, since at least Berry (1994) and Berry et al. (1995), it has been the workhorse model in the structural demand estimation literature. The ARUM relies on the single-unit purchase assumption that each consumer chooses one unit of the product that maximizes her utility given by the sum of a deterministic and a random utility terms.

Consider a linear-in-characteristics ARUM where the conditional indirect utility of consumer n choosing a product $j \in \mathcal{J}_0$ is given by

$$u_{nj} = \delta_j + \sum_{k=1}^{K_2} \sigma_k v_{nk} x_{jk}^{(2)} + \varepsilon_{nj},$$

where δ_j is given by Equation (2), σ_k, for $k = 1, \ldots, K_2$, are K_2 random coefficients, v_{nk} is a standardized random variable, $x_{jk}^{(2)}$ are K_2 exogenous characteristics, and ε_{nj} is a remaining consumer-specific valuation for product j.

Gandhi and Houde (2017) show that such a model yields inverse demands of the form

$$\sigma_j^{-1}\left(\mathbf{s}, \mathbf{x}^{(2)}; \sigma_1, \ldots, \sigma_{K_2}\right) = f\left(s_j, \{s_k, \boldsymbol{\Delta}_{j,k}^{(2)}\}_{j \neq k}; \sigma_1, \ldots, \sigma_{K_2}\right) + c, \quad j \in \mathcal{J}_0, \tag{35}$$

where $c \in \mathbb{R}$ is a market-specific constant and f is a symmetric function, with $\boldsymbol{\Delta}_{j,k}^{(2)} = \mathbf{x}_j^{(2)} - \mathbf{x}_k^{(2)}$ the vector of nonlinear characteristic differences between products j and k. They also show that inverse demands (35) exhibit symmetry, anonymity, and translation invariance in $\mathbf{x}^{(2)}$.

Economic Restrictions of the FIL Model Recall that inverse demands of the FIL model are given by $\sigma_j^{-1}(\mathbf{s}; \boldsymbol{\mu}) = \ln G_j(\mathbf{s}; \boldsymbol{\mu}) + c$, where

$$\ln G_j(\mathbf{s}; \boldsymbol{\mu}) \equiv \mu_j \ln\left(s_j\right) + \sum_{i \neq j} \mu_{ij} \ln\left(s_i + s_j\right), \quad j \in \mathcal{J}_0.$$

Then, using Assumption (i),

$$\ln G_j(\mathbf{s}; \boldsymbol{\mu}) = \ln\left(s_j\right) - \sum_{i \neq j} \mu_{ij} \ln\left(\frac{s_j}{s_i + s_j}\right), \quad j \in \mathcal{J}_0.$$

Setting $\mu_{ij} = \mu\left(\mathbf{d}_{ij}^{(2)}\left(\mathbf{x}_i^{(2)},\mathbf{x}_j^{(2)}\right)\right)$, one obtains, for all $j \in \mathcal{J}_0$,

$$\ln G_j\left(\mathbf{s},\mathbf{x}^{(2)}\right) = \ln G_j\left(\mathbf{s},\mathbf{d}^{(2)}\left(\mathbf{x}^{(2)}\right)\right) = \ln\left(s_j\right) - \sum_{i\neq j}\mu\left(\mathbf{d}_{ij}^{(2)}\left(\mathbf{x}_i^{(2)},\mathbf{x}_j^{(2)}\right)\right)\ln\left(\frac{s_j}{s_i+s_j}\right).$$

Then, $\ln G_j$ is invariant to translation in $\mathbf{x}^{(2)}$ since, for all $c \in \mathbb{R}$,

$$\ln G_j\left(\mathbf{s},\mathbf{x}^{(2)}+c\mathbf{1}_{J+1}\right) = \ln G_j\left(\mathbf{s},\mathbf{d}^{(2)}\left(\mathbf{x}^{(2)}+c\mathbf{1}_{J+1}\right)\right) = \ln G_j\left(\mathbf{s},\mathbf{d}^{(2)}\left(\mathbf{x}^{(2)}\right)\right) = \ln G_j\left(\mathbf{s},\mathbf{x}^{(2)}\right).$$

Now, rewrite $\ln G_j\left(\mathbf{s};\mathbf{x}^{(2)}\right)$ as

$$\ln G_j\left(\mathbf{s},\mathbf{x}^{(2)}\right) = \ln\left(s_j\right) - \sum_{i\in\mathcal{J}_0}\mu\left(\mathbf{d}_{ij}^{(2)}\left(\mathbf{x}_i^{(2)},\mathbf{x}_j^{(2)}\right)\right)\ln\left(\frac{s_j}{s_i+s_j}\right) + C, \quad j \in \mathcal{J}_0,$$

where the sum is over $j \in \mathcal{J}_0$, including product j, and $C = \mu_{jj}(\mathbf{0})\ln(1/2) \in \mathbb{R}$ is product-invariant. Then,

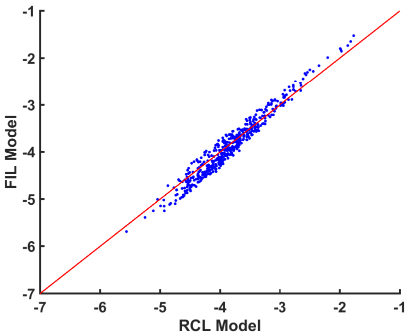
$$\ln G_j\left(\mathbf{s},\mathbf{x}^{(2)}\right) = \ln G\left((s_j,\mathbf{x}_j^{(2)}),(\mathbf{s}_{-j},\mathbf{x}_{-j}^{(2)})\right).$$

D Additional Results from Monte Carlo Simulations

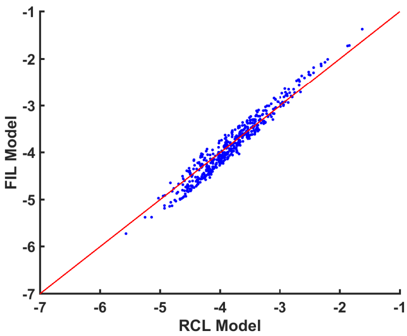
The scatter plots in Figures 1 to 5 represent the accuracy of the estimation of the FIL model when the true model is the RCL model. Each blue dot represents a price elasticity; its vertical position is the mean estimated elasticity across 500 Monte Carlo datasets and its horizontal position is the true elasticity. All market-level variables are fixed at their values in the first 20 markets of the first Monte Carlo dataset of each configuration. The red line corresponds to the 45-degree line. The scatter plots remove the outliers in terms of fit as measured by the absolute value between the estimated elasticity and its corresponding true one (to be precise, I remove the 2.5% best and the 2.5% worst fits).

Tables 4 to 6 show that the optimal instruments allow to better fit the own- and cross-price elasticities of the RCL model.

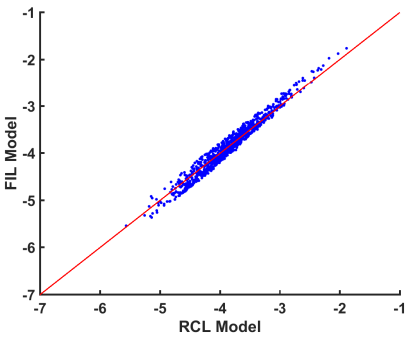
Figure 1: Results with Exogenous Prices: Own-Price Elasticities (Part 1)



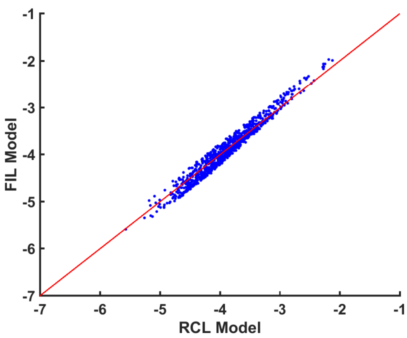
(a) $T = 20$, $J = 25$, $\sigma_x = 0.5$



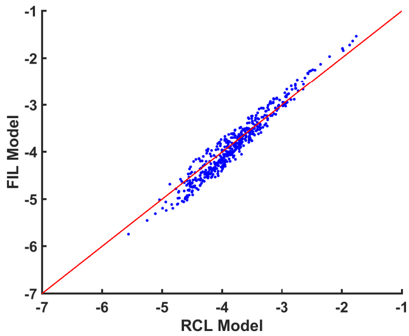
(b) $T = 50$, $J = 25$, $\sigma_x = 0.5$



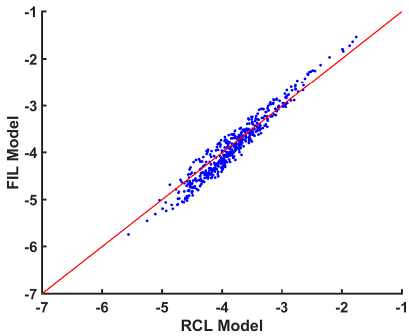
(c) $T = 20$, $J = 50$, $\sigma_x = 0.5$



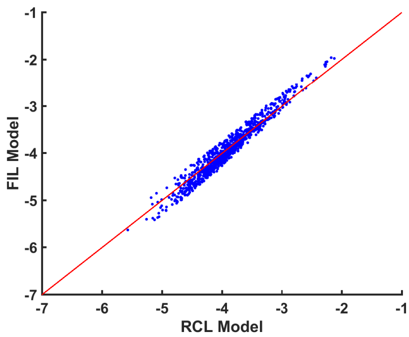
(d) $T = 50$, $J = 50$, $\sigma_x = 0.5$



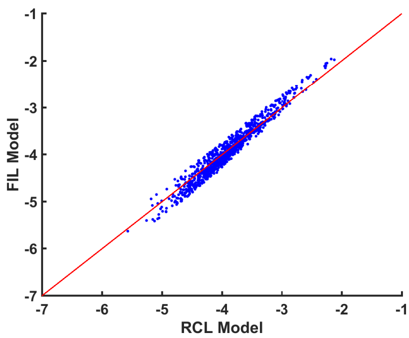
(e) $T = 20$, $J = 25$, $\sigma_x = 1$



(f) $T = 50$, $J = 25$, $\sigma_x = 1$

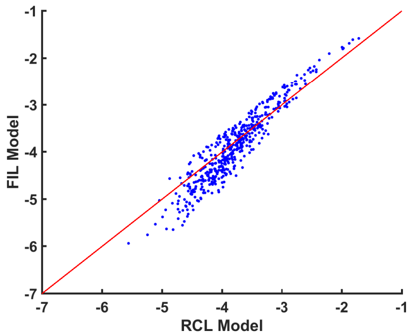


(g) $T = 20$, $J = 50$, $\sigma_x = 1$

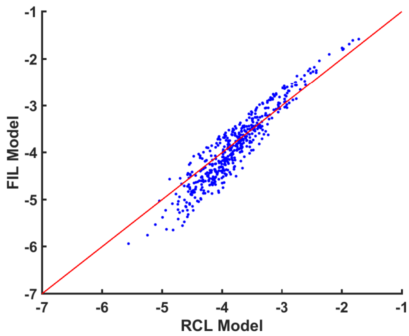


(h) $T = 50$, $J = 50$, $\sigma_x = 1$

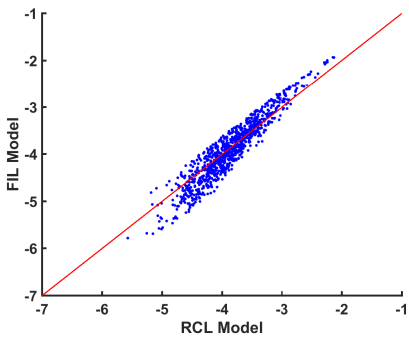
Figure 2: Results with Exogenous Prices: Own-Price Elasticities (Part 2)



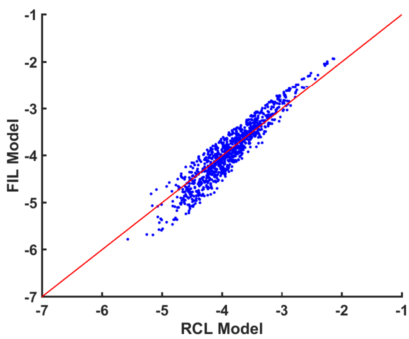
(a) $T = 20$, $J = 25$, $\sigma_x = 2$



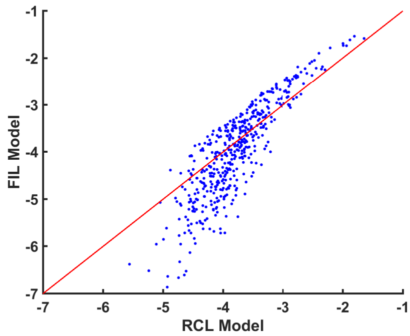
(b) $T = 50$, $J = 25$, $\sigma_x = 2$



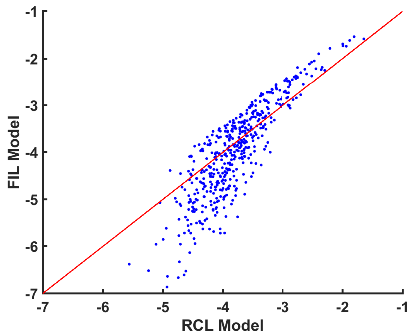
(c) $T = 20$, $J = 50$, $\sigma_x = 2$



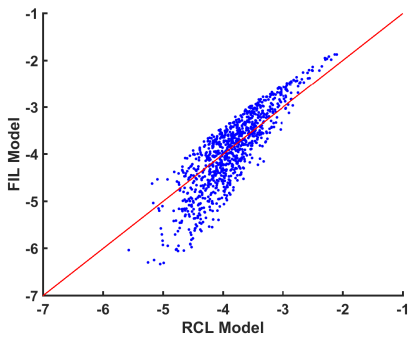
(d) $T = 50$, $J = 50$, $\sigma_x = 2$



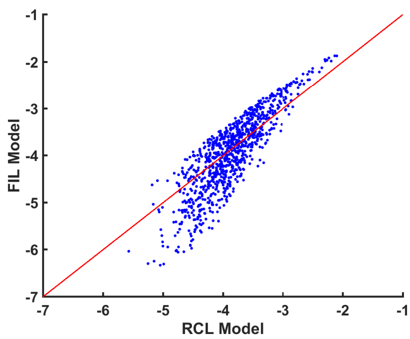
(e) $T = 20$, $J = 25$, $\sigma_x = 3$



(f) $T = 50$, $J = 25$, $\sigma_x = 3$

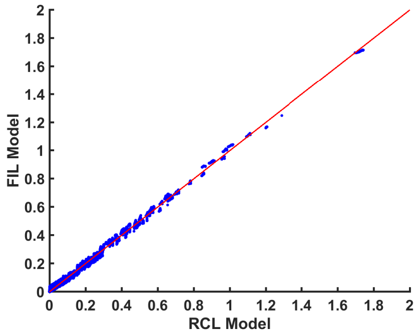


(g) $T = 20$, $J = 50$, $\sigma_x = 3$

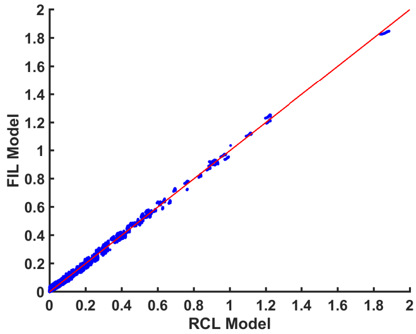


(h) $T = 50$, $J = 50$, $\sigma_x = 3$

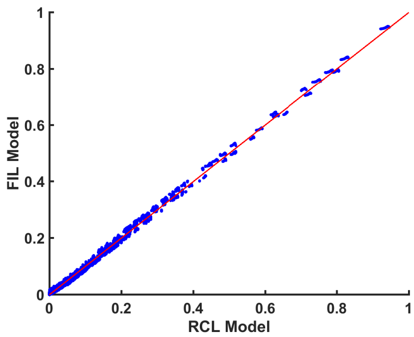
Figure 3: Results with Exogenous Prices: Cross-Price Elasticities (Part 1)



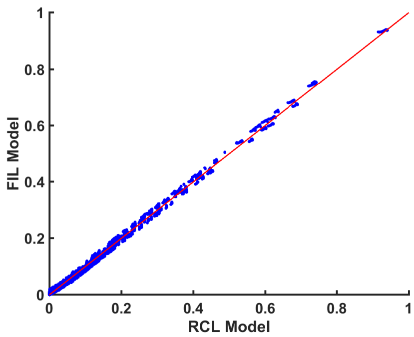
(a) $T = 20$, $J = 25$, $\sigma_x = 0.5$



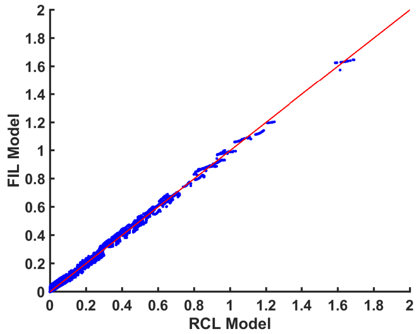
(b) $T = 50$, $J = 25$, $\sigma_x = 0.5$



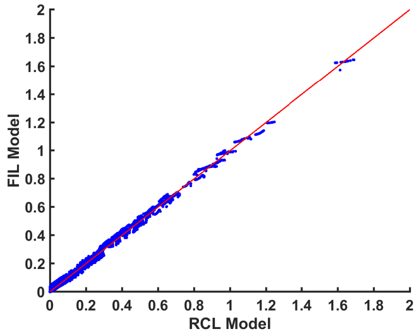
(c) $T = 20$, $J = 50$, $\sigma_x = 0.5$



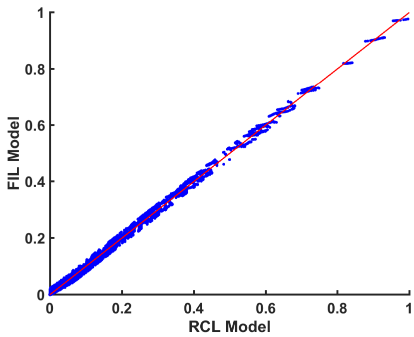
(d) $T = 50$, $J = 50$, $\sigma_x = 0.5$



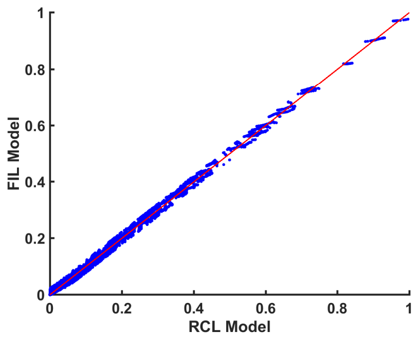
(e) $T = 20$, $J = 25$, $\sigma_x = 1$



(f) $T = 50$, $J = 25$, $\sigma_x = 1$

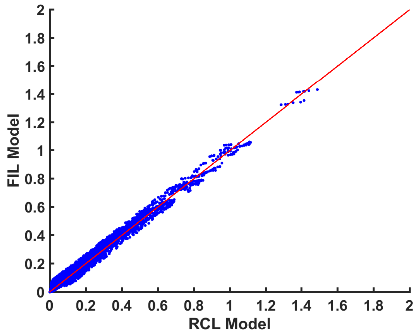


(g) $T = 20$, $J = 50$, $\sigma_x = 1$

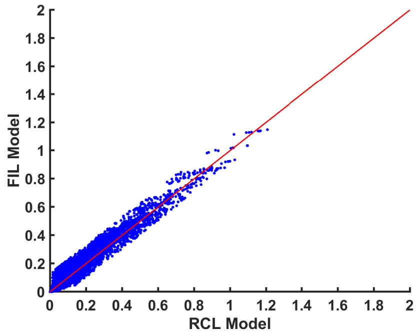


(h) $T = 50$, $J = 50$, $\sigma_x = 1$

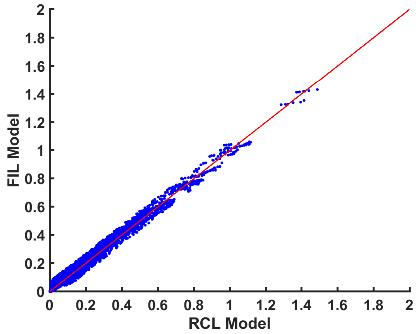
Figure 4: Results with Exogenous Prices: Cross-Price Elasticities (Part 2)



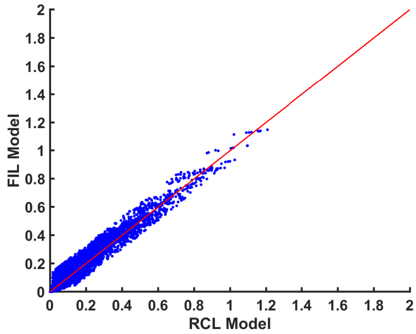
(a) $T = 20$, $J = 25$, $\sigma_x = 2$



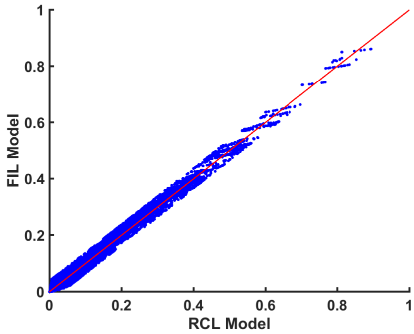
(e) $T = 20$, $J = 25$, $\sigma_x = 3$



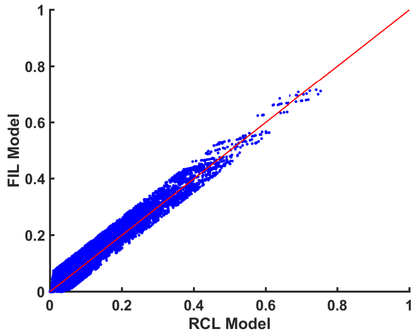
(b) $T = 50$, $J = 25$, $\sigma_x = 2$



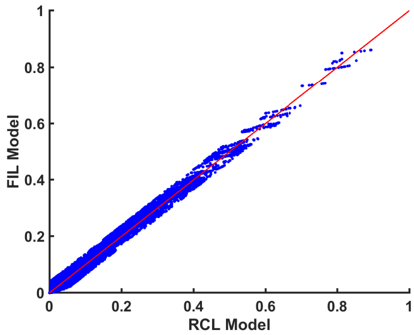
(f) $T = 50$, $J = 25$, $\sigma_x = 3$



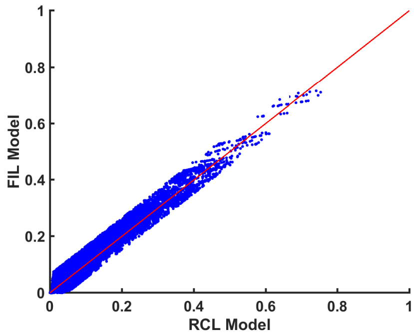
(c) $T = 20$, $J = 50$, $\sigma_x = 2$



(g) $T = 20$, $J = 50$, $\sigma_x = 3$

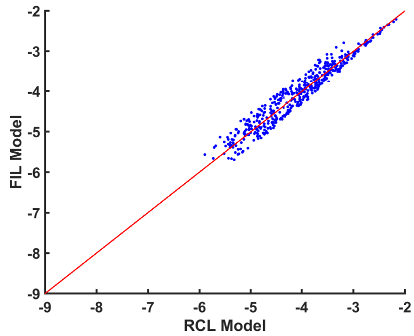


(d) $T = 50$, $J = 50$, $\sigma_x = 2$

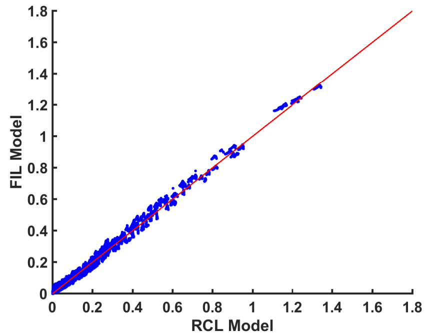


(h) $T = 50$, $J = 50$, $\sigma_x = 3$

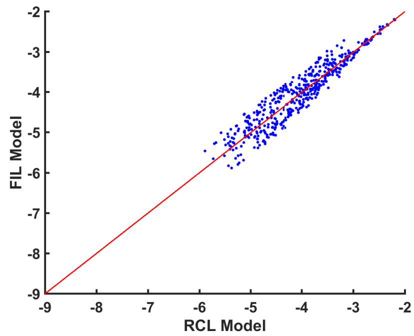
Figure 5: Results with Endogenous Prices



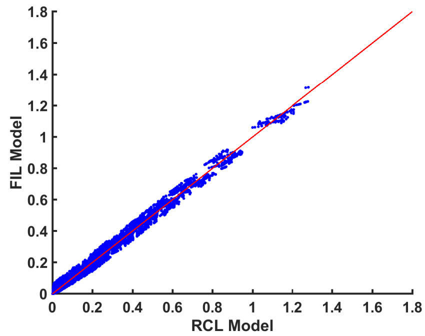
(a) Own-price elasticities, $\sigma_x = 0.5$



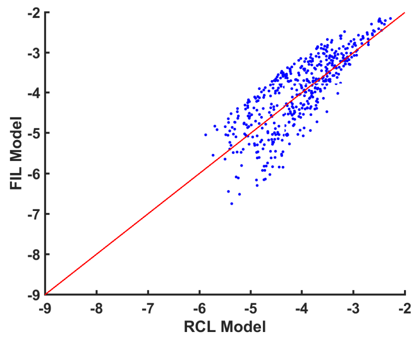
(e) Cross-price elasticities, $\sigma_x = 0.5$



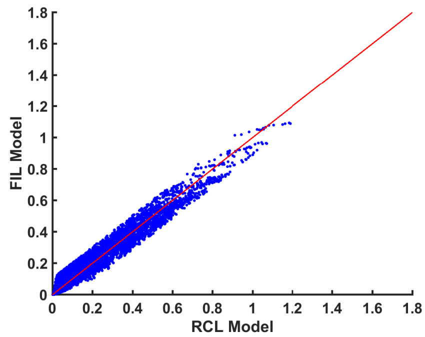
(b) Own-price elasticities, $\sigma_x = 1$



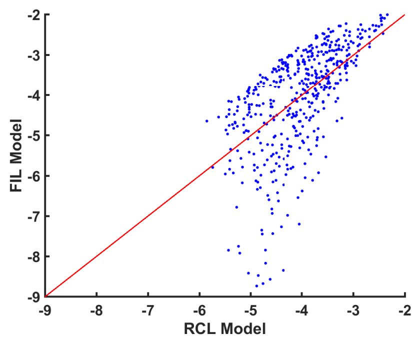
(f) Cross-price elasticities, $\sigma_x = 1$



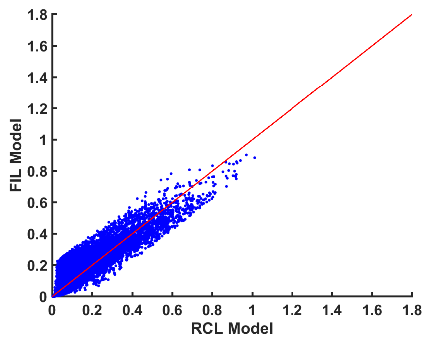
(c) Own-price elasticities, $\sigma_x = 2$



(g) Cross-price elasticities, $\sigma_x = 2$



(d) Own-price elasticities, $\sigma_x = 3$



(h) Cross-price elasticities, $\sigma_x = 3$

Table 4: Simulation Results: Own-Price Elasticities with Exogenous Prices

T	J	σ_x	Mean		Percentiles				Bias	MSE
					2.5th	97.5th	2.5th	97.5th		
			RCL	FIL	RCL		FIL			
20	25	0.5	-3.8338	-3.9197	-4.8205	-2.5582	-5.3656	-2.1735	-0.0859	0.1090
50	25	0.5	-3.8333	-3.9039	-4.8196	-2.5142	-5.3293	-2.2422	-0.0705	0.0935
20	50	0.5	-3.9021	-3.9488	-4.9011	-2.8028	-5.3510	-2.5435	-0.0467	0.0922
50	50	0.5	-3.9021	-3.9473	-4.8997	-2.7943	-5.3660	-2.5309	-0.0452	0.0847
20	25	1	-3.8314	-3.9453	-4.8191	-2.5578	-5.5691	-2.1300	-0.1139	0.1666
50	25	1	-3.8309	-3.9288	-4.8179	-2.5121	-5.4481	-2.1851	-0.0979	0.1466
20	50	1	-3.9009	-3.9628	-4.9014	-2.7982	-5.4669	-2.4962	-0.0619	0.1317
50	50	1	-3.9009	-3.9614	-4.9000	-2.7932	-5.4985	-2.4784	-0.0605	0.1241
20	25	2	-3.8201	-4.2224	-4.8144	-2.5156	-8.1138	-1.9190	-0.4023	1.4402
50	25	2	-3.8196	-4.1949	-4.8122	-2.5135	-7.3852	-1.8856	-0.3753	1.2991
20	50	2	-3.8948	-4.1436	-4.9015	-2.7790	-7.2497	-2.1976	-0.2488	0.9905
50	50	2	-3.8947	-4.1461	-4.8997	-2.7719	-7.1414	-2.1844	-0.2515	1.0192
20	25	3	-3.8018	-4.6097	-4.7905	-2.4682	-11.7957	-1.8773	-0.8079	8.5887
50	25	3	-3.8015	-4.6165	-4.8003	-2.5089	-11.7504	-1.8228	-0.8150	10.0277
20	50	3	-3.8839	-4.4182	-4.8957	-2.7562	-10.1795	-2.1065	-0.5343	5.5795
50	50	3	-3.8837	-4.4463	-4.8922	-2.7618	-9.4595	-2.0887	-0.5626	10.2463

Notes: Summary statistics across 500 Monte Carlo replications. The bias is measured by the mean error. The mean square error (MSE) measures the accuracy.

Table 5: Simulation Results: Cross-Price Elasticities with Exogenous Prices

T	J	σ_x	Mean		Percentiles				Bias	MSE
					2.5th	97.5th	2.5th	97.5th		
			RCL	FIL	RCL		FIL			
20	25	0.5	0.1515	0.1565	0.0011	0.9628	0.0006	0.8702	0.0050	0.0012
50	25	0.5	0.1520	0.1562	0.0011	0.9060	0.0007	0.8658	0.0042	0.0010
20	50	0.5	0.0752	0.0770	0.0004	0.4714	0.0003	0.4364	0.0017	0.0004
50	50	0.5	0.0752	0.0769	0.0005	0.5038	0.0003	0.4724	0.0017	0.0003
20	25	1	0.1512	0.1576	0.0016	0.9224	0.0007	0.8304	0.0064	0.0017
50	25	1	0.1517	0.1573	0.0016	0.8845	0.0008	0.8206	0.0056	0.0015
20	50	1	0.0752	0.0773	0.0007	0.4548	0.0003	0.4157	0.0021	0.0005
50	50	1	0.0752	0.0772	0.0007	0.4820	0.0003	0.4526	0.0021	0.0005
20	25	2	0.1496	0.1692	0.0037	0.8192	0.0010	0.6623	0.0196	0.0089
50	25	2	0.1501	0.1684	0.0038	0.7880	0.0011	0.6316	0.0183	0.0081
20	50	2	0.0748	0.0812	0.0017	0.4083	0.0004	0.3321	0.0065	0.0020
50	50	2	0.0748	0.0812	0.0017	0.4216	0.0005	0.3529	0.0065	0.0020
20	25	3	0.1459	0.1840	0.0062	0.7109	0.0015	0.6350	0.0381	0.0307
50	25	3	0.1463	0.1845	0.0062	0.6927	0.0015	0.6035	0.0382	0.0362
20	50	3	0.0736	0.0862	0.0029	0.3514	0.0007	0.3228	0.0126	0.0051
50	50	3	0.0736	0.0867	0.0030	0.3696	0.0009	0.3266	0.0130	0.0089

Notes: Summary statistics across 500 Monte Carlo replications. The bias is measured by the mean error. The mean square error (MSE) measures the accuracy.

Table 6: Simulation Results with Endogenous Prices

T	J	σ_x	Mean		Percentiles				Bias	MSE
					2.5th	97.5th	2.5th	97.5th		
			RCL	FIL	RCL		FIL			
Own-Price Elasticities										
50	25	0.5	-4.1208	-4.1352	-5.4317	-2.6397	-5.9828	-2.6511	-0.0145	0.2240
50	25	1	-4.1178	-4.1960	-5.4325	-2.6380	-6.6359	-2.5784	-0.0783	0.5488
50	25	2	-4.1060	-4.7652	-5.4303	-2.6084	-12.9322	-2.2572	-0.6592	14.0420
50	25	3	-4.0887	-4.7236	-5.4167	-2.5384	-14.4515	-2.1151	-0.6348	15.1808
Cross-Price Elasticities										
50	25	0.5	0.1855	0.1881	0.0038	0.8959	0.0020	0.8246	0.0026	0.0018
50	25	1	0.1843	0.1892	0.0050	0.8627	0.0020	0.7429	0.0049	0.0038
50	25	2	0.1792	0.2072	0.0087	0.7678	0.0019	0.6499	0.0281	0.0433
50	25	3	0.1720	0.2035	0.0110	0.7035	0.0028	0.6214	0.0315	0.0452

Notes: Summary statistics across 500 Monte Carlo replications. The bias is measured by the mean error. The mean square error (MSE) measures the accuracy.

E The Post-Nabisco Merger

The FIL model possesses the main features that make it appealing for merger evaluation according to Pinkse and Slade (2004): it imposes no specific restrictions on price elasticities; it is easily and fastly estimated by linear IV regression using standard computer softwares; and it can handle very large choice sets.

This appendix shows the use of the FIL model for merger simulation purposes through the study of the Post's acquisition of the Nabisco cereal line in the ready-to-eat (RTE) cereals industry that occurred in January 1993. This merger has been extensively studied in the empirical literature (see Rubinfeld (2000), Nevo (2000) and Michel and Weiergraeber (2019)).[37] Using both pre- and post-merger data, I directly estimate these price effects. Then, using pre-merger data, I estimate a FIL model and, assuming a static oligopolistic price competition model, I simulate the merging firm's price effects. Lastly, I compare the results.

E.1 Data

Datasets and Variables I use data from the Dominick's Database made available by the James M. Kilts Center, University of Chicago Booth School of Business. This is a weekly store-level scanner data, comprising information on 30 categories of packaged products at the Universal Product Code (UPC) level for all Dominick's Finer Foods (DFF) chain stores in the Chicago metropolitan area over the period 1989-1997. The data are supplemented by store-specific information, including average household size and store traffic.

For the application, I focus on the RTE cereals and use data from 60 DFF stores during the period 1992–1993, i.e., one year before and one year after the merger. I use the whole sample, i.e., both pre- and post-merger data, for the reduced form analysis, and only pre-merger data for the structural approach.

Following the prevailing literature, I aggregate UPCs into brands, where a brand is a cereal (e.g., Special K) associated to its manufacturer (e.g., Kellogg's), so that different size boxes are considered one brand. I select 45 brands from 6 national manufacturers (General Mills, Kellogg's, Nabisco, Post, Quaker and Ralston), so that they represent around 75% of each manufacturer total sales on the period. The 45 brands taken together account for around 60% of the national market (see Corts, 1996; Shum, 2004, for information on national market shares).

[37] For an industry overview, see also Corts (1996), Nevo (2001) and Backus et al. (2018).

Lastly, I group cereals into three market segments, namely Adults, Kids and Family, according to the classification provided by the website www.cerealfacts.org.

I define a product as a brand and a market as a month-store pair. I compute the market shares of the 45 brands in volume as follows. First, I select all package sizes between 10 and 32 ounces. Then, I compute the total volume sold by a brand in a market, which I divide by the potential market size to obtain the market shares. The market share of the outside option is then obtained as the difference between one and the sum of the 45 brands' market shares.

I compute the potential market size as follows. According to the USDA's Economic Research Service, per capita US consumption of RTE cereals was equal to 13.9 pounds in 1992 and 14.6 pounds in 1993. I use this information to compute the monthly per capita consumption. Then, assuming that people visit stores twice a week, I compute the total number of persons in a market as the number of households times average household size. The potential market size is thus given by the total number of persons in a market multiplied by the monthly per capita consumption of RTE cereals.

At no loss of generality, I define a serving weight as 1 ounce (i.e., 28.35g). Prices in the analysis are weighted deflated retail prices calculated as the volume-weighted average price per ounce of the UPCs that form the brand and where the deflator is the monthly Consumer Price Index for All Urban Consumers (CPI-U) in the Chicago-Naperville-Elgin area from the U.S. Bureau of Labor Statistics.[38]

I supplement the data with the fiber and sugar contents of one serve of cereals using release SR16 of the USDA Nutrient Database for Standard Reference.[39] I use information regarding the presence or not of rice, wheat, corn and oats using manufacturers' websites and different websites collecting nutritional information. I also use monthly prices for rice, wheat, sugar and corn from the website www.indexmundi.com, and for oats from the website www.macrotrends.net, which will be used to construct cost-based instruments.

Evolution of Retail Prices Before turning to the econometric analysis, observe in Figure 6 how the weighted average prices of the merging firms' and the non-

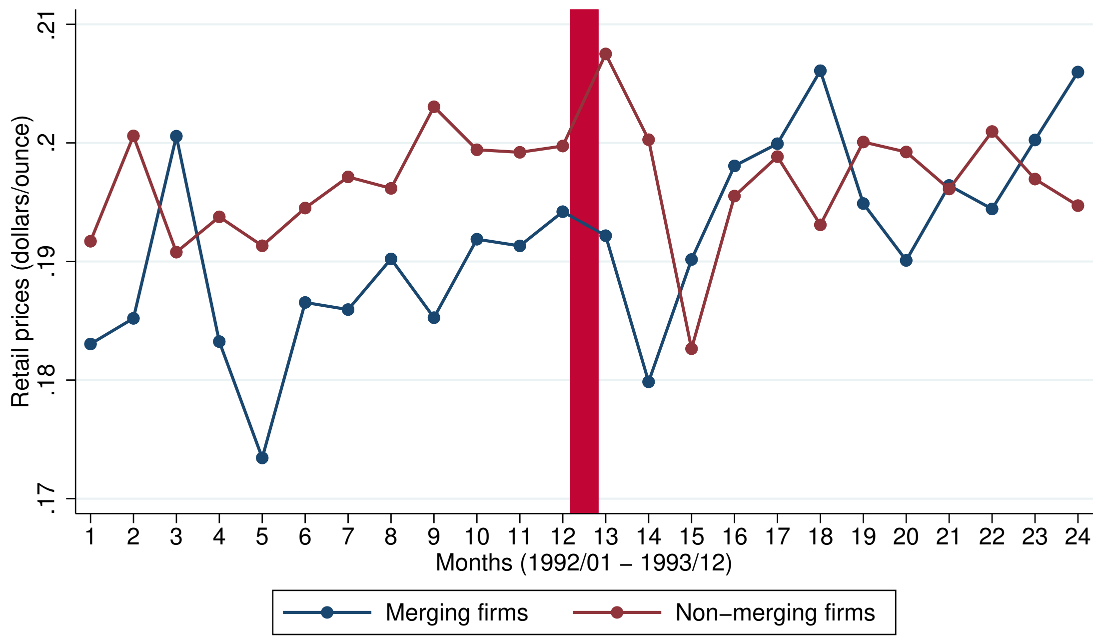
[38]The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services. Indexes are available for the U.S. and various geographic areas. Average price data for select utility, automotive fuel, and food items are also available.

[39]This dataset is made available by the United States Department of Agriculture and provides the nutrient content of more than 8,500 different foods including RTE cereals.

merging firms' separately evolved before and after the merger. It appears that both merging and non-merging firm's exhibit a slightly increasing trend in prices. In addition, before the merger, the merging firms' prices are lower than the non-merging firms' prices, while they are similar after merger.

This suggests that merging firms took advantage of the merger to increase their prices relatively to non-merging firms. However, the increasing trend can also be explained by changes in economic factors other than the merger, such that increases in price inputs. The econometric analysis of the next subsection aims at estimating the merging firms' price effect, everything else being equal.

Figure 6: Prices Evolution by Merging Status



E.2 Reduced-Form Analysis

I first estimate the price effects of the Post-Nabisco merger using both pre- and post-merger data. Based on Björnerstedt and Verboven (2016), I estimate the following regression

$$\ln(p_{jsm}) = \alpha_i + \beta_i \text{PostMerger}_i + \xi_j + \xi_s + \xi_m + \Delta\xi_{jsm},$$

where subscript i denotes a product group, and where p_{jsm} is the price of brand $j = 1, \ldots, 45$, sold in store $s = 1, \ldots, 60$ during month $m = 1, \ldots, 24$. I define product groups i at two levels: (i) merging status: $i \in \{\text{merging}, \text{non-merging}\}$; (ii) at the firm level: $i \in \{\text{General Mills}, \text{Kellogg's}, \text{Nabisco}, \text{Post}, \text{Quaker}, \text{Ralston}\}$.

The following table presents the results.

Table 7: Price Effect

	Coefficient	SE	Percent change	CI
Regression at the level of merging vs. non-merging firms				
MergingFirms × PostMerger	0.0477***	(0.00274)	4.77%	[4.23% ; 5.31%]
(1-MergingFirms) × PostMerger	0.0220***	(0.00282)	2.20%	[1.65% ; 2.76%]
(lower bound) Price effect	0.0257***	(0.00128)	2.66%	[2.40% ; 2.92%]
MergingFirms fixed effects			Yes	
Observations			64654	
R^2			0.998	
Regression at the level of the firm				
General Mills × PostMerger	0.0112***	(0.00307)	1.13%	[0.52% ; 1.74%]
Kellogg's × PostMerger	0.0283***	(0.00300)	2.87%	[2.26% ; 3.47%]
Nabisco × PostMerger	0.0556***	(0.00399)	5.72%	[4.89% ; 6.55%]
Post × PostMerger	0.0459***	(0.00275)	4.70%	[4.14% ; 5.27%]
Quaker × PostMerger	0.0117***	(0.00340)	1.18%	[0.51% ; 1.86%]
Ralston × PostMerger	0.0386***	(0.00338)	3.94%	[3.25% ; 4.63%]
Firms fixed effects			Yes	
Observations			64654	
R^2			0.998	

Notes: Regressions include fixed effects for brands, stores, and months. Robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The percentage price effects are obtained from $\exp(\beta_i) - 1$ and the corresponding standard errors using the delta method.

The structural approach of the next subsection assumes that costs did not change after the merger. To make the results consistent with the structural approach, based on Weinberg and Hosken (2013), I also estimate the following regression which controls for cost changes

$$\ln(p_{jsm}) = \alpha_i + \beta_i \text{PostMerger}_i + c_i\left(\omega_{jm}\right) + \xi_j + \xi_s + \xi_m + \Delta\xi_{jsm},$$

where $c_i\left(\omega_{jm}\right)$ is a cost function depending on price inputs which is merging status-specific. I replace $c_i\left(\omega_{jm}\right)$ by a polynomial in ω_{jm} (of orders 2 and 3, respectively). I distinguish between merging status to allow price inputs to enter the production function of differently according to the merging status. Price inputs are prices for corn, oats, rice, wheat and sugar multiplied by the content of one serving weight (in grammes for sugar and a dummy for the others).

If it is assumed that the merger has no effect on non-merging firm's prices, then the difference $\beta_{\text{merging}} - \beta_{\text{non-merging}}$ measures the merging firms' price effects. Otherwise, the difference must be viewed as a lower bound.

Table 8: Price Effect, by merging status and holding cost constant

	Coefficient	SE	Percent change	CI
Regression with a second-order polynomial cost function				
MergingFirms × PostMerger	0.0726***	(0.00355)	7.53%	[6.78% ; 8.28%]
(1-MergingFirms) × PostMerger	0.0580***	(0.00357)	5.98%	[5.23% ; 6.72%]
(lower bound) Price effect	0.0145***	(0.00150)	1.55%	[1.24% ; 1.87%]
Observations			64654	
R^2			0.998	
Regression with a third-order polynomial cost function				
MergingFirms × PostMerger	0.0444***	(0.00409)	4.54%	[3.71% ; 5.38%]
(1-MergingFirms) × PostMerger	0.0245***	(0.00408)	2.48%	[1.66% ; 3.30%]
(lower bound) Price effect	0.0199***	(0.00178)	2.06%	[1.70% ; 2.42%]
Observations			64654	
R^2			0.998	

Notes: Regressions include fixed effects for brands, stores, and months. Robust standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The percentage price effects are obtained from $\exp(\beta_i) - 1$ and the corresponding standard errors using the delta method.

E.3 Structural Approach using the FIL Model

The structural approach uses the structural model of demand and supply presented in subsection 5.1.

Demand Estimation I estimate the FIL model given by Equation (19) where $\mathbf{x}^{(1)}$ includes a constant as well as dummies for brands, stores and months, and where $\mathbf{x}^{(2)}$ includes the fiber and sugar contents of the cereals – product characteristics $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are invariant across markets, this is the reason for which I omit notation for markets t.

I estimate the model following the two-step procedure described in Subsection 5.3. In the first step, I estimate the FIL model using a first-order Bernstein polynomial (with 2×2 parameters), where instruments are given by

$$\mathrm{IV}_{jt} = \left\{ \mathbf{x}_j, \hat{p}_{jt}, \left(\sum_{i \neq j} b_{k_1,1} b_{k_2,2} \hat{p}_{it} \right)_{k_1=\{0,1\};k_2=\{0,1\}} \right\},$$

where $\hat{p}_{jt}$ is the predicted value of the linear regression of prices p_{jt} on $\mathbf{x}_j^{(1)}$, $\left(\mathbf{x}_j^{(1)}\right)^2$, $\sum_{i \neq j} \left(\mathbf{x}_i^{(2)} - \mathbf{x}_j^{(2)}\right)$ and ω_{jt}. Cost shifters are input prices (for sugar, corn, oats, rice and wheat) multiplied by their content in one serving weight (in grammes for sugar and a dummy for the others).[40]

In the second step, I use a second-order Bernstein polynomial (with 3×3 parameters), with the optimal instruments of Chamberlain (1987). The Sanderson and Windmeijer (2016)'s F-statistics of the 10 first-stage regressions are far higher than 10, indicating that instruments are not weak.

The estimated median own-price elasticities are in line with the literature. They range from -6.181 to -1.461 with an average of -3.362. Regarding the cross-price elasticities, they range from -0.692 to 0.669, with an average of 0.038. 30% of the pairs of products are complements (see Iaria and Wang, 2019, who find a large amount of complementarity in the RTE cereals industry). Note, however, that confidence intervals should be computed to determine whether or not the small values for some of the cross-price elasticities are well significantly different from zero.

Merger Simulation Consider a static oligopolistic price competition model between the six manufacturers. Assuming that a pure-strategy Nash equilibrium exists and using the associated first-order conditions, I compute marginal costs and margins. The model predicts negative marginal costs only for less than 0.130% of the observations. Note that I use retail prices to simulate a merger between manufacturers, while ignoring the retailers (DFF) – I abstract from the vertical relationship between retailers and manufacturers. The median marginal cost implied by the model is 0.13 dollars per serving, which, in line with Michel and Weiergraeber (2019), implies median margin equal to 33 cents.

Given the demand estimates and the predicted marginal costs, I simulate the merger by only changing the merging firms' product ownership. I compute the

[40] Bernstein polynomials used in this application have lower orders than those used in the simulations. This is because with higher orders, instruments become weak. A future version of this paper will consider Bernstein polynomials with higher orders.

post-merger prices using the approximate solution of the first-order conditions (see Hausman et al., 1994; Nevo, 1997), and then I compute the merging firms' price effect as

$$\sum_{j\in\mathcal{M}}\sum_{s=1}^{60}\sum_{m=1}^{12} w_{jsm}\frac{p_{jsm}^{\text{post}}-p_{jsm}}{p_{jsm}},$$

with $\sum_{j,s,m} w_{jsm} = 1$, where $\mathcal{M}$ is the set of brands sold by the merging firms, where w_{jsm} are premerger volume shares, and where p_{jsm}^{post} are the simulated post-merger prices.

E.4 Comparison

Assuming that the merger has no effect on non-merging firm's prices, the reduced-form approach leads to an estimate of the merging firm's price effect of 1.55% when the cost function is approximated by a second-order polynomial in input prices and of 2.06% when it is approximated by a third-order polynomial. These results must be taken with caution, especially because there are cost factors that are not considered in the analysis, such as cost of packaging, distribution, advertising, etc.

The structural approach when the demand model is the FIL model and the supply model is a static price competition model leads to an estimate of the merging firm's price effect of 2.03%, thereby indicating that the FIL model predicts a merger effect on retail prices in line with that found in the reduced-form approach. However, several specifications and robustness checks should be run before getting a reliable conclusion.

Bibliography

Allen, R. and J. Rehbeck (2019): "Identification With Additively Separable Heterogeneity," *Econometrica*, 87, 1021–1054.

Anderson, S. P., A. De Palma, and J. F. Thisse (1992): *Discrete choice theory of product differentiation*, Cambridge, MA: MIT press.

Andrews, I., J. Stock, and L. Sun (2018): "Weak instruments in IV regression: Theory and practice," *Working paper*.

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

Backus, M., C. Conlon, and M. Sinkinson (2018): "Common ownership and competition in the ready-to-eat cereal industry," *Working paper*.

Barnett, W. A. (1983): "Definitions of "second order approximation" and of "flexible functional form"," *Economics Letters*, 12, 31–35.

——— (1985): "The minflex-Laurent translog flexible functional form," *Journal of Econometrics*, 30, 33–44.

Berry, S., A. Eizenberg, and J. Waldfogel (2016): "Optimal product variety in radio markets," *The RAND Journal of Economics*, 47, 463–497.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

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

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

——— (1999): "Voluntary export restraints on automobiles: Evaluating a trade policy," *American Economic Review*, 89, 400–430.

Berry, S. and A. Pakes (2001): "Additional information for:Comments on Alternative models of demand for automobiles by Charlotte Wojcik," *Economics Letters*, 74, 43–51.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 25, 242–262.

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

Björnerstedt, J. and F. Verboven (2016): "Does merger simulation work? Evidence from the Swedish analgesics market," *American Economic Journal: Applied Economics*, 8, 125–64.

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

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

Bulow, J. I. and P. Pfleiderer (1983): "A note on the effect of cost changes on prices," *Journal of political Economy*, 91, 182–185.

Chak, P. M., N. Madras, and B. Smith (2005): "Semi-nonparametric estimation with Bernstein polynomials," *Economics Letters*, 89, 153–156.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

Chu, C. (1989): "A Paired Combinational Logit Model for Travel Demand Analysis," in *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, Vol. 4, Western Periodicals, Ventura, CA, 295–309.

Compiani, G. (2019): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," *Unpublished*.

Conlon, C. and J. Gortmaker (2018): “Best practices for differentiated products demand estimation with pyblp,” *Working paper*.

Corts, K. (1996): “The Ready-to-eat Breakfast Cereal Industry in 1994 (A),” *Harvard Business School*.

Cournot, A.-A. (1838): *Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot*, chez L. Hachette.

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.

Davis, P. J. (1975): *Interpolation and approximation*, Courier Corporation.

Diewert, W. (1974): “Application of Duality Theory,” *Frontiers of Quantitative Economics*, 2.

Doraszelski, U. and A. Pakes (2007): “A framework for applied dynamic analysis in IO,” *Handbook of industrial organization*, 3, 1887–1966.

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.

Ershov, D., J.-W. Laliberté, and S. Orr (2018): “Mergers in a Model with Complementarity,” *Available at SSRN: https://ssrn.com/abstract=3295232*.

Fosgerau, M., J. Monardo, and A. de Palma (2019): “The Inverse Product Differentiation Logit Model,” *Available at SSRN: https://ssrn.com/abstract=3141041*.

Gandhi, A. and J.-F. Houde (2017): “Measuring substitution patterns in differentiated products industries,” *Unpublished*.

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

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

Griffith, R., M. O’Connell, and K. Smith (2019): “Tax design in the alcohol market,” *Journal of Public Economics*, 172, 20–35.

Grigolon, L. (2018): "Blurred boundaries: a flexible approach for segmentation applied to the car market," *Working paper*.

Hanemann, W. M. (1984): "Discrete/continuous models of consumer demand," *Econometrica: Journal of the Econometric Society*, 541–561.

Hausman, J., G. Leonard, and J. D. Zona (1994): "Competitive analysis with differenciated products," *Annales d'Economie et de Statistique*, 159–180.

Hurwicz, L. (1966): "On the structural form of interdependent systems," in *Studies in Logic and the Foundations of Mathematics*, Elsevier, vol. 44, 232–239.

Iaria, A. and A. Wang (2019): "Identification and Estimation of Demand for Bundles: The Case of the RTE Cereal Industry," *Unpublished*.

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

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.

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

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

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 15, 447–470.

Michel, C. and S. Weiergraeber (2019): "Estimating Industry Conduct in Differentiated Products Markets," *Working paper*.

Miller, N. H., M. Remer, C. Ryan, and G. Sheu (2016): "Pass-Through and the Prediction of Merger Price Effects," *The Journal of Industrial Economics*, 64, 683–709.

Monardo, J. (2019): "Shape Restrictions for Structural Demand Estimation: Beyond the ARUM," *Unpublished*.

Mrázová, M. and J. P. Neary (2017): "Not so demanding: Demand structure and firm behavior," *American Economic Review*, 107, 3835–74.

Nevo, A. (1997): "Mergers with Differentiated Products: The Case of Ready-to-Eat Cereal," .

——— (2000): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 31, 395–421.

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

Newey, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica: Journal of the Econometric Society*, 58, 809–837.

Newman, P. (2008): "Substitutes and complements," *The New Palgrave Dictionary of Economics: Volume 1–8*, 6434–6438.

Ogaki, M. (1990): "The indirect and direct substitution effects," *The American Economic Review*, 80, 1271–1275.

Petrin, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of political Economy*, 110, 705–729.

Pinkse, J. and M. E. Slade (1998): "Contracting in space: An application of spatial statistics to discrete-choice models," *Journal of Econometrics*, 85, 125–154.

——— (2004): "Mergers, brand competition, and the price of a pint," *European Economic Review*, 48, 617–643.

Pinkse, J., M. E. Slade, and C. Brett (2002): "Spatial price competition: a semiparametric approach," *Econometrica*, 70, 1111–1153.

Reiss, P. C. (2016): "Just How Sensitive are Instrumental Variable Estimates?" *Foundations and Trends in Accounting*, 10, 204–237.

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

Rojas, C. (2008): "Price competition in US brewing," *The Journal of Industrial Economics*, 56, 1–31.

RUBINFELD, D. L. (2000): "Market definition with differentiated products: the Post/Nabisco cereal Merger," *Antitrust LJ*, 68, 163.

SALANIÉ, B. AND F. A. WOLAK (2019): "Fast," robust", and approximately correct: estimating mixed demand systems," *National Bureau of Economic Research*.

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

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

SHUM, M. (2004): "Does advertising overcome brand loyalty? Evidence from the breakfast-cereals market," *Journal of Economics & Management Strategy*, 13, 241–272.

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

SLADE, M. E. (2004): "Market power and joint dominance in UK brewing," *The Journal of Industrial Economics*, 52, 133–163.

VERBOVEN, F. (1996): "International price discrimination in the European car market," *The RAND Journal of Economics*, 240–268.

WEINBERG, M. C. AND D. HOSKEN (2013): "Evidence on the accuracy of merger simulations," *Review of Economics and Statistics*, 95, 1584–1600.

WEYL, E. G. AND M. FABINGER (2013): "Pass-through as an economic tool: Principles of incidence under imperfect competition," *Journal of Political Economy*, 121, 528–583.

WOJCIK, C. (2000): "Alternative models of demand for automobiles," *Economics Letters*, 68, 113–118.

Chapter 3

Shape Restrictions for Demand Estimation

1 Introduction

Structural estimation of demand models for differentiated products allows to better understand consumers' behavior and to study many economic questions of interest (e.g., mergers, new products, trade policy, cost pass-through, etc.). Ideally, one would like demand models to impose limited shape restrictions while being consistent with an underlying structural model of heterogeneous and utility-maximizing consumers.

Most papers that structurally estimate demands assume an additive random utility model (ARUM) from which the demand function is derived and then estimated using the method developed by Berry (1994). The ARUM is ubiquitously used due to its ability to model the behavior of heterogeneous consumers in the presence of many differentiated products in a tractable and parsimonious way.[1] Berry (1994)'s method is used to estimate demand models for differentiated products while handling endogeneity issues, due to the presence of structural error terms that represent product characteristics that are unobserved by the modeller but observed and valued by consumers and firms.

This paper is motivated by two observations. First, the ARUM imposes shape

[1]The ARUM relies on the single-unit purchase assumption that each consumer chooses one unit of the product that maximizes her utility given by the sum of a deterministic and a random utility terms. It has been widely applied in empirical industrial organization (see Berry, 1994; Berry, Levinsohn, and Pakes, 1995) and in other fields to answer many questions (see e.g., Table 1 in Berry and Haile, 2016).

restrictions on its associated demand function, known as the Daly-Zachary conditions due to Daly and Zachary (1979).[2] In particular, they rule out complementarity as defined by a negative cross-price derivative of demand, which may be undesirable in applications where complementarity is likely to occur and to qualitatively affect results (e.g, mergers, bundling, products' entry).[3] Second, Berry (1994)'s method targets the inverse demand function, rather than the demand function or the underlying utility function. This is because it starts from the equation that equates the predicted to the observed market shares where the structural errors enter nonlinearly, thereby preventing the use of standard instrumental variables (IV) methods to deal with endogeneity issues. The method thus suggests to invert it to obtain inverse demand equations where the structural errors enter linearly and to use them as a basis for estimation. However, since these inverse demand equations have generally no closed form expression, estimation requires numerical inversion and non-linear optimization, which can be painful and time-consuming.[4]

Fosgerau et al. (2019b) contrasts with the standard practice by building novel inverse demand models, called generalized inverse logit (GIL) models, which allow for complementarity and which are estimated by IV regression using Berry (1994)'s method. Interestingly, Fosgerau et al. (2019b) show that any GIL model can be derived from a model of heterogeneous and utility-maximizing consumers, called perturbed utility model (PUM) and studied by Allen and Rehbeck (2019a), see arrow "Lemma 1" in Figure 1.[5]

The main goal of this paper is to show the converse, namely that, under mild conditions, any PUM yields an inverse demand function that is a GIL model. The proof of this result involves two intermediate results that are of independent interest. First, any PUM yields a demand function satisfying a slight variant of the Daly-Zachary conditions, referred to as the modified Daly-Zachary (MDZ)

[2]In this paper, I use interchangeably the terms "demands", "choice probabilities", and "market shares".

[3]I use the standard definition of complementarity (substitutability), i.e. a negative (positive) compensated cross-price derivative of demand. Since I rule out income effect, complementarity (substitutability) is the only source of the negative (positive) cross-price derivative of demand. See Samuelson (1974) for a discussion on the different ways of defining complementarity.

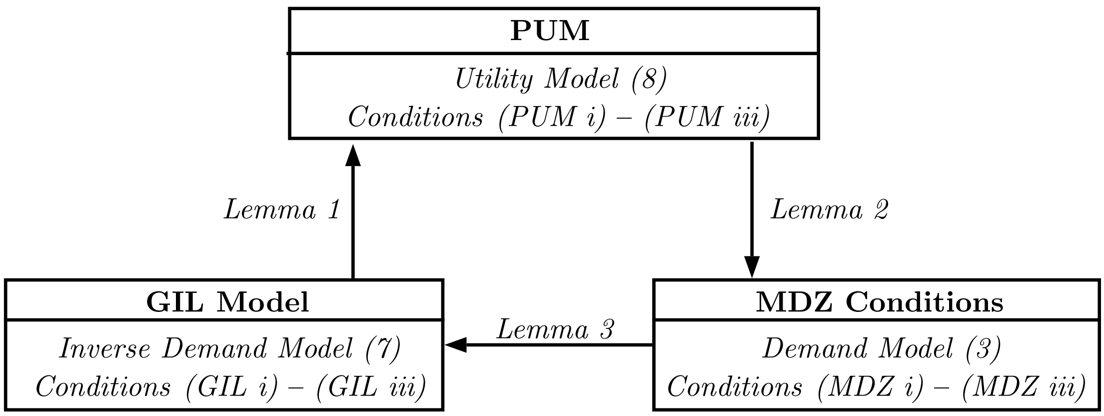
[4]To my knowledge, the logit and nested logit models are the only ARUM with a closed-form inverse demand function.

[5]The PUM assumes that each consumer chooses a probability distribution over products so as to maximize her utility given by the sum of an expected utility term and a non-linear, deterministic function of probabilities. The PUM extends the ARUM to allow for richer substitution patterns.

conditions and allowing for complementarity, see arrow "Lemma 2" in Figure 1. Second, any demand function satisfying the MDZ conditions has an inverse demand function that is a GIL model, see arrow "Lemma 3" in Figure 1. Overall, by equivalence relations, PUM, GIL models and demands satisfying the MDZ conditions are observationally equivalent.

The remainder of the paper is organized as follows. Section 2 introduces the setting used in the paper. Section 3 defines the PUM and the GIL model. Section 4 establishes the main result of the paper, namely that the shape restrictions that the GIL model imposes on the inverse demand function are necessary and sufficient for consistency with PUM maximization, and sketches its proof. Section 5 provides the proofs of the results of the paper. Section 6 concludes.

Figure 1: Relations between PUM, GIL Models and MDZ Conditions



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

Let $\mathcal{J} = \{0,\ldots,J\}$. A vector $\mathbf{s} \in \mathbb{R}^{J+1}$ refers to $\mathbf{s} \equiv \left(s_0,\ldots,s_J\right)^\top \in \mathbb{R}^{J+1}$.

Δ_J denotes the J-dimensional unit simplex: $\Delta_J \equiv \left\{\mathbf{s} \in [0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$, $\mathrm{int}\left(\Delta_J\right) \equiv \left\{\mathbf{s} \in (0,\infty)^{J+1} : \sum_{j\in\mathcal{J}} s_j = 1\right\}$ is its interior, and $\mathrm{bd}(\Delta_J) \equiv \Delta_J \backslash \mathrm{int}\left(\Delta_J\right)$ is its boundary.

Let $CS : \mathbb{R}^{J+1} \to \mathbb{R}$ be a function. Then, $\nabla_{\delta} CS(\delta)$, with entries j given by $\frac{\partial CS(\delta)}{\partial \delta_j}$, denotes its gradient with respect to the vector δ.

Let $\mathbf{G} = \left(G_0,\ldots,G_J\right) : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a vector function composed of functions $G_j : \mathbb{R}^{J+1} \to \mathbb{R}$. The Jacobian matrix $\mathbf{J}^{\mathbf{s}}_{\mathbf{G}}(\bar{\mathbf{s}})$ of $\mathbf{G}$ with respect to $\mathbf{s}$ at $\bar{\mathbf{s}}$ has entries ij given by $\frac{\partial G_i(\bar{\mathbf{s}})}{\partial s_j}$.

A univariate function $\mathbb{R} \to \mathbb{R}$ applied to a vector is a coordinate-wise application of the function, e.g., $\ln(\mathbf{s}) = \left(\ln(s_0), \ldots, \ln\left(s_J\right)\right)$. $\mathbf{1}_J = (1, \ldots, 1)^\intercal \in \mathbb{R}^J$ is a vector consisting of ones.

$|\mathbf{s}|$ denotes the 1-norm of the vector $\mathbf{s}$: $|\mathbf{s}| = \sum_{j \in \mathcal{J}} |s_j|$. $\boldsymbol{\delta}_{-j}$ denotes the vector $\boldsymbol{\delta}$ after deleting its j-th entry: $\boldsymbol{\delta}_{-j} \equiv (\delta_0 \ldots \delta_{j-1}, \delta_{j+1} \ldots \delta_J)$.

2 Setting

Consider a population of consumers choosing from a choice set $\mathcal{J} = \{0, \ldots, J\}$ of $J+1$ differentiated products, where products $j = 1, \ldots, J$ are the inside products and product $j = 0$ is the outside good. Let $p_j \in \mathbb{R}$ be the price of product j and $\mathbf{x}_j \in \mathbb{R}^K$ be the vector of K observed characteristics of product j. Following the structural demand estimation literature (Berry, 1994; Berry et al., 1995), let $\xi_j \in \mathbb{R}$ be the j-product unobserved characteristics term. The ξ_j's represent the structural error terms of the demand model in the sense that they are assumed to be observed by consumers and firms but not by the modeller.

Following Berry and Haile (2014), define for each product $j \in \mathcal{J}$ a linear index

$$\delta_j = \mathbf{x}_j \boldsymbol{\beta} - \alpha p_j + \xi_j, \quad j \in \mathcal{J}, \tag{1}$$

where $\boldsymbol{\beta}$ captures the consumers' taste for characteristics $\left(\mathbf{x}_0, \ldots, \mathbf{x}_J\right)$ and $\alpha > 0$ denotes the consumers' marginal utility of income; and normalize the indexes of each inside product relative to that of the outside good by setting $\delta_0 = 0$, so that $\boldsymbol{\delta} \in \mathbb{R}_0^{J+1} \equiv \{\boldsymbol{\delta} \in \mathbb{R}^{J+1} : \delta_0 = 0\}$.[6]

Consider the demand system

$$\boldsymbol{\sigma} = \left(\sigma_0, \ldots, \sigma_J\right) : \mathbb{R}_0^{J+1} \to \operatorname{int}\left(\Delta_J\right),$$

where the function $\boldsymbol{\sigma}$ gives the vector $\mathbf{s}$ of nonzero observed market shares as a function of the vector of product indexes $\boldsymbol{\delta}$,

$$\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta}; \boldsymbol{\theta}_2), \tag{2}$$

known up to some parameters $\boldsymbol{\theta}_2$ to be estimated. Demands (2) rule out income effect (since they are independent of income); the implicit assumption behind

[6] Alternatively, one can normalize $\boldsymbol{\delta}$ such that $\mathbb{R}_0^{J+1} = \{\boldsymbol{\delta} \in \mathbb{R}^{J+1} : \sum_{j \in \mathcal{J}} \delta_j = 0\}$.

being that preferences are quasi-linear and that income is sufficiently high that not all income is spent on the $J+1$ differentiated products (see e.g., Nocke and Schutz, 2017).

Demand estimation based on the system (2) that equates the predicted market shares to the observed market shares is complicated by the fact that the error terms $\boldsymbol{\xi}$ enter in a nonlinear way, thereby preventing the use of standard regression techniques. Berry (1994) suggests inverting system to obtain an inverse demand system in which the error terms $\boldsymbol{\xi}$ enter linearly and to use it as a basis for estimation. Indeed, if $\boldsymbol{\sigma}$ is invertible in $\boldsymbol{\delta}$, then there exists an inverse demand $\boldsymbol{\sigma}^{-1} : \text{int}(\Delta_J) \to \mathbb{R}_0^{J+1}$ given by

$$\boldsymbol{\sigma}^{-1}(\mathbf{s};\boldsymbol{\theta}_2) = \boldsymbol{\delta},$$

in which the structural error terms ξ_j enter linearly as a function of the data (i.e., market shares, prices and product characteristics) and parameters α, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}_2$ to be estimated,[7]

$$\xi_j = \sigma_j^{-1}(\mathbf{s};\boldsymbol{\theta}_2) - \mathbf{x}_j\boldsymbol{\beta} + \alpha p_j, \quad j \in \mathcal{J}. \tag{3}$$

Invertibility in $\boldsymbol{\delta}$ of the system of market shares (2) is thus crucial for identification and estimation. Berry et al. (2013) show that their "connected substitutes" structure is sufficient for demand invertibility.[8] Proposition 2 provides a result of invertibility for demands that can accommodate some patterns of substitution that are not allowed in Berry et al. (2013)'s setting, including complementarity as defined by a negative cross-price derivative of demand.

Product characteristics are typically assumed to be exogenous (i.e., that they are uncorrelated with the structural error terms). However, prices and market shares in the right-hand side of Equation (3) are likely to be endogenous. Then, provided that there exists appropriate instruments $\mathbf{z}$ for prices and market shares, one can estimate demands (2) based on the following conditional moment restrictions

$$\mathbb{E}\left[\xi_j|\mathbf{z}\right] = 0, \quad j \in \mathcal{J},$$

where ξ_j is given by Equation (3).

[7] To ease exposition, I hereafter omit notation for parameters $\boldsymbol{\theta}_2$.

[8] The connected substitutes structure requires that (i) products be weak substitutes, i.e., everything else equal, an increase in δ_i weakly decreases demand σ_j for all other products and (ii) the "connected strict substitution" condition hold, i.e., there is sufficient strict substitution between products to treat them in one demand system.

3 Models

Most papers that structurally estimate demands assume an additive random utility model (ARUM), derive the associated demand function, and invert it to obtain the inverse demand equations (3). However, the ARUM rules out complementarity by assumption and, in general, yields an inverse demand function that has no closed-form expression, so that one needs to resort to numerical inversion and non-linear optimization for estimation.

Fosgerau et al. (2019b)'s approach constrasts with this practice, obtaining Equation (3) by directly developing novel (closed-form) inverse demand models, called generalized inverse logit (GIL), which allow for complementarity and just require IV regression for estimation.[9]

Fosgerau et al. (2019b) shows that any GIL model is consistent with a model of heterogeneous and utility-maximizing consumers, called perturbed utility model (PUM) and studied by Allen and Rehbeck (2019a).[10] The goal of this paper is to show the converse, namely that the inverse demand function yield by any PUM is a GIL model, to establish that the conditions that the GIL model imposes on the inverse demand function are necessary and sufficient for consistency with PUM maximization.

3.1 The Generalized Inverse Logit (GIL) Model

The GIL model is an inverse demand model of the form

$$\sigma_j^{-1}(\mathbf{s}) = \ln G_j(\mathbf{s}) + c = \delta_j, \quad j \in \mathcal{J}, \tag{4}$$

where δ_j is given by Equation (1), where c is equal, up to an additive constant, to the consumer surplus and where the function $\ln \mathbf{G} \equiv (\ln G_0, \ldots, \ln G_J)$ satisfies the following conditions.

Conditions GIL The function $\ln \mathbf{G}$ satisfies the following conditions.

(GIL i) $\mathbf{G} : [0,\infty)^{J+1} \to (0,\infty)^{J+1}$ is continuously differentiable and homogeneous of degree one on $\operatorname{int}\left(\Delta_J\right)$[11]

[9]Compiani (2019) adopts a similar approach. Using Berry and Haile (2014)'s setting, he proposes to non-parametrically estimate inverse demand models.

[10]See also McFadden and Fosgerau (2012) and Fudenberg et al. (2015).

[11]The function $\mathbf{G}$ is homogeneous of degree one if, for all $\lambda > 0$, $\mathbf{G}(\lambda \mathbf{s}) = \lambda \mathbf{G}(\mathbf{s})$.

(GIL ii) $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ is positive definite and symmetric on $\operatorname{int}\left(\Delta_J\right)$[12]

(GIL iii) $|\ln \mathbf{G}(\mathbf{s})|$ approaches infinity as $\mathbf{s}$ approaches $\operatorname{bd}(\Delta_J)$.

This form of inverse demand generalizes the inverse demand obtained from the logit model through the presence of the function $\mathbf{G}$, and as shown by Fosgerau et al. (2019b), allows for complementarity.

3.2 The Perturbed Utility Model (PUM)

The PUM is a utility model that admits a representative consumer choosing a vector $\mathbf{s} \in \Delta_J$ of market shares of the differentiated products so as to maximize her quasi-linear utility defined as the sum of an expected utility component and a perturbation function $(-\Omega)$

$$u(\mathbf{s}) = \sum_{j \in \mathcal{J}} \delta_j s_j - \Omega(\mathbf{s}), \tag{5}$$

where δ_j is given by Equation (1) and where the perturbation function $(-\Omega)$ is a deterministic function of $\mathbf{s}$ that satisfies the following conditions.

Conditions PUM The function $\Omega : [0, \infty)^{J+1} \to \mathbb{R} \cup \{+\infty\}$ satisfies the following conditions.[13]

(PUM i) Ω is finite for $\mathbf{s} \in \Delta_J$ and infinite otherwise

(PUM ii) Ω is twice continuously differentiable and strictly convex on $\operatorname{int}(\Delta_J)$.

(PUM iii) $|\nabla_{\mathbf{s}} \Omega(\mathbf{s})|$ approaches infinity as $\mathbf{s}$ approaches $\operatorname{bd}(\Delta_J)$.

Conditions *(PUM i)* – *(PUM iii)* imply that the argmax of the utility maximization program is unique and that this maximizer is interior, i.e., $\boldsymbol{\sigma} \in \operatorname{int}(\Delta_J)$. Demands are then given by

$$\boldsymbol{\sigma}(\boldsymbol{\delta}) = \operatorname*{argmax}_{\mathbf{s} \in \Delta_J} \{u(\mathbf{s})\}, \tag{6}$$

[12] $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ is positive definite on $\operatorname{int}\left(\Delta_J\right)$ if $\mathbf{s}^\intercal \mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}} \mathbf{s} > 0$, for all $\mathbf{s} \in \operatorname{int}\left(\Delta_J\right)$.

[13] Conditions *(PUM i)* to *(PUM iii)* imply that $(\operatorname{int}(\Delta_J), \Omega)$ is a convex function of Legendre type. See Appendix A.1. See, amongst others, Hofbauer and Sandholm (2002), Galichon and Salanié (2015), Matějka and McKay (2015), Chiong et al. (2016), Allen and Rehbeck (2019a), Fosgerau et al. (2019a) for economics papers that use convex analysis.

where u is given by Equation (5).

As shown by Allen and Rehbeck (2019a), this form of utility can be derived, after an aggregation across consumers, from a model of heterogeneous and utility-maximizing consumers called perturbed utility model (PUM). Indeed, in a PUM, the utility that each consumer derives from choosing a vector $\mathbf{s} \in \Delta_J$ is the sum of an expected utility $\sum_{j \in \mathcal{J}} \delta_j s_j$ term and a disturbance function $(-\tilde{\Omega})$,

$$u(\mathbf{s}; \varepsilon) = \sum_{j \in \mathcal{J}} \delta_j s_j - \tilde{\Omega}(\mathbf{s}, \varepsilon),$$

where ε capture consumer heterogeneity, so that the utility (5) can be obtained after integrating out the distribution of ε.[14] [15]

The PUM is consistent with deliberate stochastic choice at the consumer level (Fudenberg et al., 2015; Allen and Rehbeck, 2019b), with stochastic choice due to rational inattention (Matějka and McKay, 2015; Fosgerau et al., 2019a) or due to costly optimization (Mattsson and Weibull, 2002), and with taste for variety.[16]

As noted by Allen and Rehbeck (2019a), the ARUM is a special case of the PUM, where $\tilde{\Omega}(\mathbf{s}, \varepsilon) = -\sum_{j \in \mathcal{J}} s_j \varepsilon_j$, meaning that the class of PUM is strictly larger than the class of ARUM (see also Hofbauer and Sandholm, 2002; Fosgerau et al., 2019a,b).

The PUM allows for complementarity (Allen and Rehbeck, 2018), which is ruled out by the ARUM. Complementarity in a PUM can be due to a variety of consumer behavior, including taste for variety and one-stop shopping, and can also be the manifestation of the attraction effect or the compromise effect (Rieskamp et al., 2006).

[14]See Theorem 1 in Allen and Rehbeck (2019a) for the conditions under which this aggregation holds and for the relationships between the functions $\tilde{\Omega}$ and Ω and the distribution of ε. See also Allen and Rehbeck (2019b).

[15]This specification assumes that the only source of consumer heterogeneity in preferences is due to the vector of random utility terms ε whose distribution is parametrized by the vector $\boldsymbol{\theta}_2$. This assumption rules out observed heterogeneity in preferences related to observed individual characteristics as well as unobserved heterogeneity in preferences through random coefficients on price and product characteristics. This implies that the probability that each consumer chooses product $j \in \mathcal{J}$ coincides with the market share of that product. For this reason, I omit notation for consumers.

[16]See also Machina (1985), Clark (1990), Agranov and Ortoleva (2017) and Cerreia-Vioglio et al. (2019).

4 Conditions for Consistency with PUM Maximization

This section establishes the main result of the paper. It also sketches its proof, which is illustrated in Figure 1. The findings of this section are illustrated through the logit example in Appendix C. For the purpose of this paper, the modified Daly-Zachary (MDZ) conditions are introduced. They are demand shape restrictions extending the Daly-Zachary conditions due to Daly and Zachary (1979).

The following proposition states that the conditions that the GIL models impose on the inverse demand function are necessary and sufficient for consistency with PUM maximization.

Proposition 1. Any GIL model is consistent with a PUM, i.e., it can be derived from the maximization of a utility function of the form of Equation (5). Conversely, any PUM yields an inverse demand function that is a GIL model.

The proof of this proposition uses three lemmas, stated in this section and formally proved in the next section, which link the PUM, the GIL model and the MDZ conditions.

The first lemma, shown in Fosgerau et al. (2019b), establishes that any GIL model is consistent with a PUM. Specifically, the GIL model (4) is consistent with a consumer choosing a vector $\mathbf{s} \in \Delta_J$ of market shares of the differentiated products so as to maximize her quasi-linear utility

$$u(\mathbf{s}) = \sum_{j \in \mathcal{J}} \delta_j s_j - \sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s}), \tag{7}$$

which is a PUM where the perturbation $-\Omega$ given by

$$\Omega(\mathbf{s}) = \begin{cases} \sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s}) & \text{if } \mathbf{s} \in \Delta_J, \\ +\infty & \text{otherwise,} \end{cases}$$

satisfies Conditions *(PUM i)* – *(PUM iii)*. This result is summarized as follows.

Lemma 1. The GIL model (4) is consistent is the PUM (7).

The second lemma shows that any PUM yields a demand function satisfying the modified Daly-Zachary (MDZ) conditions given as follows.

Conditions MDZ The demand function $\boldsymbol{\sigma}$ satisfies, for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$ and for all $j \in \mathcal{J}$,

(MDZ i) $\sigma_j(\boldsymbol{\delta}) > 0$ and $\sum_{k \in \mathcal{J}} \sigma_k(\boldsymbol{\delta}) = 1$ (positive unit demand)

(MDZ ii) $\sigma_j(\boldsymbol{\delta}) = \sigma_j(\boldsymbol{\delta} + c\mathbf{1})$ for all $c \in \mathbb{R}$ (translation invariance)

(MDZ iii) $\lim\limits_{\delta_j \to \infty} \sigma_j(\boldsymbol{\delta}) = 1$ and $\lim\limits_{\delta_j \to \infty} \sigma_i(\boldsymbol{\delta}) = 0$ with $\delta_i < \infty$, $i \neq j$ (boundary conditions)

(MDZ iv) $\mathbf{J}^{\delta}_{\sigma}$ is positive definite on $\mathbb{R}^{J+1}_0$ (Slutsky positive definiteness)

(MDZ v) $\mathbf{J}^{\delta}_{\sigma}$ is symmetric (Slutsky symmetry).

The MDZ conditions represent a slight variant of the Daly-Zachary conditions, due to Daly and Zachary (1979), which are necessary and sufficient for consistency with ARUM maximization (see Appendix B). As discussed below, they are weaker than the Daly-Zachary conditions. The following lemma thus helps understanding how the PUM extends the ARUM.

Lemma 2. The demand function (6) yield by the PUM (5) satisfies Conditions *(MDZ i) – (MDZ v)*.

The third lemma shows that any demand function satisfying the MDZ conditions has an inverse function that is a GIL model. For this purpose, let, for all $j \in \mathcal{J}$,

$$\sigma_j(\boldsymbol{\delta}) \equiv \tilde{\sigma}_j\left(e^{\boldsymbol{\delta}}\right) = s_j,$$

which, by translation invariance, is equivalent to

$$\sigma_j(\boldsymbol{\delta} + c\mathbf{1}) = \tilde{\sigma}_j\left(e^{\boldsymbol{\delta}}\right) = s_j. \tag{8}$$

By the inverse function theorem, Slutsky positive definiteness implies that $\boldsymbol{\sigma}$ is one-to-one on $\mathbb{R}^{J+1}_0$, so that one can invert Equation (8) to obtain

$$\sigma_j^{-1}(\mathbf{s}) = \ln G_j(\mathbf{s}) + c = \delta_j, \tag{9}$$

where $\ln \tilde{\sigma}_j^{-1}(\mathbf{s}) \equiv \ln G_j(\mathbf{s})$, which is of the form of Equation (4).

Using tools from convex analysis, one can further show (see Section 5) that c is equal, up to an additive constant, to the consumer surplus and that $\ln \mathbf{G} = (\ln G_0, \ldots, \ln G_J)$ satisfies Conditions *(GIL i) – (GIL iii)*.

The following lemma shows how the MDZ conditions imposed on a demand function translate to its inverse.[17]

Lemma 3. Assume that the demand function $\boldsymbol{\sigma}$ satisfies Conditions *(MDZ i)* – *(MDZ v)*. Then, its inverse $\boldsymbol{\sigma}^{-1}$ is a GIL model.

Lemma 3 shows that any demand function satisfying the MDZ conditions has an inverse demand that is a GIL model, thereby establishing that the GIL structure is sufficient to recover all demand functions satisfying the MDZ conditions. Since the MDZ conditions are weaker than the Daly-Zachary conditions, this also implies that the class of GIL models is strictly larger than the class of ARUM choice probabilities defined in Appendix B. This has already been shown by Fosgerau et al. (2019b) but using a different line of proof.

Discussion on the MDZ Conditions The MDZ conditions are written in terms of product indexes. However, since these latter are linear in prices (see Equation (1)), they can easily be re-written in terms of prices.[18] Except otherwise stated, in this paper, Slutsky matrix refers to the matrix of demand derivatives with respect to product indexes $\mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}}$ (and not with respect to prices $\mathbf{J}_{\boldsymbol{\sigma}}^{\mathbf{p}}$).

The MDZ conditions have an economic content. Condition *(MDZ i)* states that all products are chosen with non-null probability and that demands sum to one. Condition *(MDZ ii)* states that demands are invariant to translation in product indexes $\boldsymbol{\delta}$, which implies that only differences in product indexes determine demands, not their absolute values. Since product indexes are linear in prices, this also implies that demands depend on price differences. As a consequence, demand models satisfying this condition require a normalization on $\boldsymbol{\delta}$ for identification, i.e., for there to be a unique vector $\boldsymbol{\delta}$ that rationalizes the vector of observed market shares $\mathbf{s}$ (see Proposition 2 below). This also implies that one can restrict $\boldsymbol{\sigma}$ to the set $\mathbb{R}_0^{J+1}$ without loss of generality. Condition *(MDZ iii)* means that all consumers choose product $j \in \mathcal{J}$ with certainty when that product becomes infinitely attractive (i.e., when δ_j goes to infinity), with the others remaining finitely attractive (i.e., with δ_i finite for all $i \neq j$). Conditions *(MDZ iv)*

[17] Compiani (2019) discusses shape restrictions on inverse demands satisfying the "connected substitutes" structure (see his Appendix D). In particular, he shows that it implies that the Slutsky matrix $\mathbf{J}_{\boldsymbol{\sigma}}^{\mathbf{p}}$ is an M-matrix and satisfies weak column diagonal dominance, so that its inverse is an inverse M-matrix and is weakly diagonally dominant of its row entries.

[18] For example, Slutsky symmetry $\frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial \delta_i} = \frac{\partial \sigma_i(\boldsymbol{\delta})}{\partial \delta_j}$ is equivalent to $\frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial p_i} = \frac{\partial \sigma_i(\boldsymbol{\delta})}{\partial p_j}$. Likewise, Slutsky negative definiteness $\mathbf{z}^\intercal \mathbf{J}_{\boldsymbol{\sigma}}^{\mathbf{p}} \mathbf{z} < 0$ for all $\mathbf{z} \in \mathbb{R}^{J+1}$ is equivalent to $\mathbf{z}^\intercal \mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}} \mathbf{z} > 0$ for all $\mathbf{z} \in \mathbb{R}^{J+1}$.

and *(MDZ v)* are key for rationalizability of demands, i.e., for them to be consistent with utility maximization.

The MDZ conditions relax the Daly-Zachary by replacing positivity with Slutsky positive definiteness. Positivity is a technical condition that has no economic content and that is generally hard to verify in practice. It is required for the implied distribution of the random utility components to have a positive density. On top of that, it rules out complementarity, as defined by a negative cross-price derivatives. By contrast, Slutsky positive definiteness accommodates substitution patterns that go beyond those obtained from ARUM. In particular, it allows to obtain demands that combine substitutes and complements.

Slutsky symmetry and positive definiteness are well known conditions. Slutsky symmetry is already part of the Daly-Zachary conditions. In addition, as shown by Hofbauer and Sandholm (2002), any demand function satisfying Conditions *(MDZ i)* and *(MDZ v)* and only allowing for substitutes admits a Slutsky matrix that is positive definite on $\mathbb{R}_0^{J+1}$. Then, in ARUM, Slutsky positive definiteness is not restrictive with respect to Slutsky positive semi-definiteness. These conditions are also well known conditions in continuous-choice models (see e.g., Hurwicz, 1971; Lewbel, 2001; Nocke and Schutz, 2017). As shown by Nocke and Schutz (2017), continuously differentiable demands are consistent with quasi-linear utility maximization if and only if $\mathbf{J}_\sigma^{\mathbf{p}}$ is symmetric and negative semi-definite, i.e., if and only if $\mathbf{J}_\sigma^{\delta}$ is symmetric and positive semi-definite.

Invertibility of Demand As a by-product, the following proposition, which is of independent interest, shows invertibility of GIL models.

Proposition 2. Assume that $\ln \mathbf{G}$ satisfies Conditions *(GIL i) – (GIL iii)*.

(i) It follows that $\ln \mathbf{G}$ is invertible on $\text{int}(\Delta_J)$.

(ii) Consider any vector of market shares $\mathbf{s} \in \text{int}\left(\Delta_J\right)$. Then, there exists a unique $\boldsymbol{\delta} \in \mathbb{R}_0^{J+1}$ such that

$$\delta_j = \sigma_j^{-1}(\mathbf{s}) = \ln G_j(\mathbf{s}) + c, \quad j \in \mathcal{J}, \tag{10}$$

or equivalently,

$$s_j = \sigma_j(\boldsymbol{\delta}) = \frac{H_j\left(e^{\boldsymbol{\delta}}\right)}{\sum_{k \in \mathcal{J}} H_k\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J}, \tag{11}$$

where $H_j \equiv G_j^{-1}$.

Part (i) of Proposition 2 uses convex analysis to show the invertibility of any function $\ln \mathbf{G}$ satisfying Conditions *(GIL i) – (GIL iii)*.

Equation (10) describes an inverse demand, i.e., a mapping from market shares to product indexes, up to a normalizing additive constant due to Condition *(MDZ ii)*. Part (ii) of Proposition 2 establishes existence and uniqueness of its inverse mapping (11) from product indexes to market shares (i.e., the demand function). This proposition thus states that the vector of product indexes $\boldsymbol{\delta}$ is identified up to an additive constant c from the vector of observed nonzero market shares $\mathbf{s}$ by the relation (10), where c is fixed by normalizing δ_0 to zero.

Proposition 2 supplements, and in some cases extends, other results on demand invertibility. Demands satisfying the MDZ Conditions allow for complementarity. This implies that Proposition 2 extends Berry (1994)'s invertibility result, which assumes substitutability. It also supplements Berry et al. (2013) to allow for complementarity. Their result holds for demands satisfying the "connected substitutes" structure, which rules out complementarity as defined by a negative cross-price derivative of demand but, by contrast to Proposition 2, does not require demand differentiability. Complementarity violates the first condition of the connected substitutes structure, which, however, accommodates some form of complementarity (see Example 1 in Berry et al. (2013) and Section 4.3. in Compiani (2019)). A related invertibility result can be found in Fosgerau et al. (2019b) (see Proposition 1), who use a different line of proof.

Using Equation (10), different inverse demand models can be obtained from different specifications of the function $\mathbf{G} = \left(G_0, \ldots, G_J\right)$. Equation (10) can thus serve as general-purpose specification tool: by specifying a function $\mathbf{G}$, the modeller determines the way products interact in utility (7) or in demand (11), and thus the type of relationship between them. In other words, Equation (10) gives a general method for parametrizing the cross-price elasticities of demands.

Combining Equations (10) for product 0 and product j, with the normalization $\delta_0 = 0$, one obtains the following inverse demand equations to be estimated

$$\xi_j = -\ln\left(\frac{G_j(\mathbf{s})}{G_0(\mathbf{s})}\right) + \mathbf{x}_j \boldsymbol{\beta} - \alpha p_j, \quad j = 1, \ldots, J. \tag{12}$$

Then, after parametrizing $\mathbf{G}$ with parameters $\boldsymbol{\theta}_2$, Equations (12) can be estimated using standard IV regression techniques. In addition, if $\ln \mathbf{G}$ is linear in parameters $\boldsymbol{\theta}_2$, then one just requires linear IV regression for estimation; this is the case of the logit and nested logit models (Berry, 1994; Verboven, 1996), of the inverse product differentiation logit model (Fosgerau et al., 2019b) and of the flexible inverse logit model (Monardo, 2019). Alternatively, Equation (12) could be estimated non-parametrically, imposing shape restrictions on **G** given by Conditions *(GIL i)* – *(GIL iii)*.

5 Proofs

This section provides the proofs of Lemmas 2 and 3 and of Proposition 2.

5.1 Proof of Lemma 2

Condition *(MDZ i)* By definition, $\boldsymbol{\sigma}$ satisfies $\sum_{j\in\mathcal{J}}\sigma_j(\boldsymbol{\delta}) = 1$ with $\sigma_j(\boldsymbol{\delta}) \geq 0$ for all $j \in \mathcal{J}$.

In the PUM, the representative consumer solves

$$\max_{\mathbf{s}\in\Delta_J}\left\{\sum_{j\in\mathcal{J}}\delta_j s_j - \Omega(\mathbf{s})\right\}. \tag{13}$$

The corresponding Lagrangian is given by

$$\mathcal{L}(\mathbf{s},\lambda,\lambda_0,\ldots,\lambda_J) = \sum_{j\in\mathcal{J}}\delta_j s_j - \Omega(\mathbf{s}) + \lambda\left(1 - \sum_{j\in\mathcal{J}} s_j\right) + \sum_{j\in\mathcal{J}}\lambda_j s_j,$$

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

This utility maximization program leads to $\sum_{j\in\mathcal{J}} s_j = 1$ and the following first-order conditions

$$\delta_j - \frac{\partial\Omega(\mathbf{s})}{\partial s_j} - \lambda + \lambda_j = 0, \quad j \in \mathcal{J}.$$

Note that the objective function of the utility maximization program (13) is continuous on the compact set Δ_J. Then, by the Weierstrass theorem, the utility maximizing program has a solution, either in the interior or on the boundary of Δ_J.

Condition *(GIL iii)* ensures that this solution is interior. To see this, note that for $\mathbf{s} \in \mathrm{bd}(\Delta_J)$, $|\nabla_{\mathbf{s}}\Omega(\mathbf{s})| = +\infty$. This implies that $\mathbf{s} \in \mathrm{bd}(\Delta_J)$ cannot solve the first-order conditions, and, in turn, that the solution is interior, i.e., satisfies $\sum_{j\in\mathcal{J}}\sigma_j(\boldsymbol{\delta}) = 1$ with $\sigma_j(\boldsymbol{\delta}) > 0$ for all $j \in \mathcal{J}$.

Lastly, the strict concavity of Ω, which implies the strict concavity of the objective function, ensures that the solution is unique.

Condition *(MDZ ii)* is implied by the quasi-linearity of the utility function as follows

$$\begin{aligned}\underset{\mathbf{s}\in\Delta_J}{\operatorname{argmax}}\left\{\sum_{j\in\mathcal{J}}\left(\delta_j+c\right)s_j-\Omega(\mathbf{s})\right\}&=\underset{\mathbf{s}\in\Delta_J}{\operatorname{argmax}}\left\{\sum_{j\in\mathcal{J}}\delta_j s_j-\Omega(\mathbf{s})+c\right\},\\&=\underset{\mathbf{s}\in\Delta_J}{\operatorname{argmax}}\left\{\sum_{j\in\mathcal{J}}\delta_j s_j-\Omega(\mathbf{s})\right\}.\end{aligned}$$

(see also Allen and Rehbeck, 2019a, p.1031). This implies that the vector $\boldsymbol{\delta}$ is identified up to an additive constant, which, in turn, requires a normalization. In the remainder of this proof, I consider, without loss of generality, that $\boldsymbol{\delta}$ belongs to $\mathbb{R}_0^{J+1}\equiv\{\boldsymbol{\delta}\in\mathbb{R}^{J+1}:\delta_0=0\}$.

Condition *(MDZ iii)* Assume that δ_j tends towards infinity, while the other δ_i, $i\neq j$, remain finite. Then, since $\Omega(\mathbf{s})$ is finite on Δ_J, this implies that s_j tends towards its upper bound, i.e., one, and, then, that the other s_i, $i\neq j$, tend towards zero (since $\mathbf{s}\in\operatorname{int}(\Delta_J)$).

Condition *(MDZ iv)* One can use Lemma 1 of Allen and Rehbeck (2019a), since the PUM considered in the present paper satisfies their Assumption 2. Then, the PUM satisfies Roy's identity, i.e., $\nabla_{\boldsymbol{\delta}}CS(\boldsymbol{\delta})=\boldsymbol{\sigma}(\boldsymbol{\delta})$. In addition, as shown in the proof of Proposition 2, $(\operatorname{int}(\Delta_J),\Omega)$ is a convex function of Legendre type. Then, by Proposition 4, $(\mathbb{R}_0^{J+1},CS)$ is also a convex function of Legendre type, which implies that CS is strictly convex on $\mathbb{R}_0^{J+1}$, and thus has a Hessian that is positive definite. Overall, this implies that Condition *(MDZ iv)* is satisfied.

Condition *(MDZ v)* is implied by Lemma 2 of Allen and Rehbeck (2019a), since the PUM considered in the present paper satisfies their Assumption 2.

5.2 Proof of Lemma 3

Note first, by Condition *(MDZ ii)*, that one can consider without loss of generality the restriction of $\boldsymbol{\sigma}$ and CS to $\mathbb{R}_0^{J+1}$. Then, by the inverse function theorem, Condition *(MDZ iv)* implies that $\boldsymbol{\sigma}$ is one-to-one on $\mathbb{R}_0^{J+1}$. Set

$$\sigma_j(\boldsymbol{\delta}) = \overline{\sigma}_j(\boldsymbol{\delta}) = s_j, \quad j \in \mathcal{J}. \tag{14}$$

By Condition *(MDZ ii)*, this is equivalent to

$$\sigma_j(\boldsymbol{\delta} + c\mathbf{1}) = \overline{\sigma}_j(\boldsymbol{\delta}) = s_j, \quad j \in \mathcal{J},$$

which, after inverting, gives

$$\sigma_j^{-1}(\mathbf{s}) = \overline{\sigma}_j^{-1}(\mathbf{s}) + c = \delta_j, \quad j \in \mathcal{J}.$$

Now, set $\overline{\sigma}_j(\boldsymbol{\delta}) = \tilde{\sigma}_j\left(e^{\boldsymbol{\delta}}\right)$, for all $j \in \mathcal{J}$. Inverting gives $\overline{\sigma}_j^{-1}(\mathbf{s}) = \ln \tilde{\sigma}_j^{-1}(\mathbf{s}) \equiv \ln G_j(\mathbf{s})$, so that one obtains the requested result

$$\sigma_j^{-1}(\mathbf{s}) = \ln G_j(\mathbf{s}) + c = \delta_j, \quad j \in \mathcal{J}. \tag{15}$$

Condition *(MDZ v)* implies that $\boldsymbol{\sigma}$ admits a function $\tilde{CS} : \mathbb{R}_0^{J+1} \to \mathbb{R}$ such that $\nabla_{\boldsymbol{\delta}} \tilde{CS} = \boldsymbol{\sigma}$. By Roy's identity, this function is, up to an additive constant, the consumer surplus, $\tilde{CS} = CS$. In addition, Condition *(MDZ iv)* implies that CS is differentiable and convex on $\mathbb{R}_0^{J+1}$. Then, by Proposition 3, $\nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) = \boldsymbol{\sigma}(\boldsymbol{\delta}) = \mathbf{s}$ implies that

$$\sum_{j \in \mathcal{J}} \delta_j s_j = \Omega(\mathbf{s}) + CS(\boldsymbol{\delta}), \tag{16}$$

where $\Omega : \mathrm{int}(\Delta_J) \to \mathbb{R}$ is the convex conjugate of CS.

Now, note that c is in utility terms, that is, $c = c(\boldsymbol{\delta})$. Then, multiplying Equations (15) by s_j and summing over $j \in \mathcal{J}$ yields

$$\sum_{j \in \mathcal{J}} \delta_j s_j = \sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s}) + c(\boldsymbol{\delta}). \tag{17}$$

Comparing the Equations (16) and (17) leads to $c(\boldsymbol{\delta}) = CS(\boldsymbol{\delta})$.

Condition *(GIL i)* Using Equation (14), Conditions *(MDZ iv)* and *(MDZ v)* imply that $\mathbf{J}_{\overline{\sigma}}^{\delta}$ is positive definite and symmetric on $\mathbb{R}_0^{J+1}$.

Recall that $\overline{\boldsymbol{\sigma}} \equiv \tilde{\boldsymbol{\sigma}} \circ \exp$ is one-to-one; then set $\mathbf{s} = \tilde{\boldsymbol{\sigma}} \circ \exp(\boldsymbol{\delta})$, so that $\boldsymbol{\delta} =$

$\ln \tilde{\sigma}^{-1}(\mathbf{s}) \equiv \ln \mathbf{G}(\mathbf{s})$. Using the inverse function theorem, one obtains

$$\mathbf{J}_{\tilde{\sigma}}^{\delta}(\delta) = \mathbf{J}_{\tilde{\sigma}\circ\exp}^{\delta}(\delta) = \left[\mathbf{J}_{(\tilde{\sigma}\circ\exp)^{-1}}^{\mathbf{s}}(\tilde{\sigma}\circ\exp(\delta))\right]^{-1} = \left[\mathbf{J}_{\ln\tilde{\sigma}^{-1}}^{\mathbf{s}}(\mathbf{s})\right]^{-1} = \left[\mathbf{J}_{\ln\mathbf{G}}^{\mathbf{s}}(\mathbf{s})\right]^{-1},$$

i.e., $\mathbf{J}_{\ln\mathbf{G}}^{\mathbf{s}}(\mathbf{s}) = \left[\mathbf{J}_{\tilde{\sigma}}^{\delta}(\delta)\right]^{-1}$, which implies that $\mathbf{J}_{\ln\mathbf{G}}^{\mathbf{s}}$ is symmetric and positive definite on $\text{int}(\Delta_J)$.

Condition *(GIL ii)* Using Equation (17), Proposition 3 implies that $\mathbf{s} \in \text{int}(\Delta_J)$ maximizes

$$\sum_{j\in\mathcal{J}} \delta_j s_j - \sum_{j\in\mathcal{J}} s_j \ln G_j(\mathbf{s}),$$

which yields the first-order conditions $\sum_{j\in\mathcal{J}} s_j = 1$ together with

$$\delta_j = \ln G_j(\mathbf{s}) + \sum_{k\in\mathcal{J}} s_k \frac{\partial \ln G_k(\mathbf{s})}{\partial s_j} + \lambda, \quad j \in \mathcal{J}.$$

By symmetry of $\mathbf{J}_{\ln\mathbf{G}}^{\mathbf{s}}$, this yields

$$\delta_j = \ln G_j(\mathbf{s}) + \sum_{k\in\mathcal{J}} s_k \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} + \lambda, \quad j \in \mathcal{J}.$$

However, recall from Equation (15),

$$\delta_j = \ln G_j(\mathbf{s}) + c,$$

where c is common across products. This means that the quantity

$$\sum_{k\in\mathcal{J}} s_k \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k}$$

is independent of $\mathbf{s}$ and common across products. This also implies that it can be set, e.g., equal to $K \in \mathbb{R}$, so that

$$\sum_{k\in\mathcal{J}} s_k \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} = K,$$

which, by Lemma 4, means that $\mathbf{G}$ is homogeneous of degree K. Then,

$$\delta_j = \ln G_j(\mathbf{s}) + K + \lambda, \quad j \in \mathcal{J},$$

which, after inverting, yields

$$s_j = H_j\left(e^{\delta-(K+\lambda)}\right) = e^{-(K+\lambda)/K} H_j\left(e^{\delta}\right),$$

where the second equality uses that $\mathbf{H}$ is homogeneous of degree $1/K$, its inverse $\mathbf{G}$ being homogeneous of degree K.

Using that demands sum to one yields $s_j = \frac{H_j(e^{\delta})}{\sum_{k\in\mathcal{J}} H_k(e^{\delta})}$ and $\ln\left(\sum_{k\in\mathcal{J}} H_k\left(e^{\delta}\right)\right) = (K+\lambda)/K$, which implies that $CS(\delta) = K\ln\left(\sum_{k\in\mathcal{J}} H_k\left(e^{\delta}\right)\right)$. However, demands must satisfy Roy's identity, i.e., $\sigma_j(\boldsymbol{\delta}) = \frac{\partial CS(\boldsymbol{\delta})}{\partial \delta_j}$, which implies that $K = 1$.

Condition *(GIL iii)* By Proposition 4, $(\nabla_{\mathbf{s}}\Omega)^{-1}(\boldsymbol{\delta}) = \nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) = \boldsymbol{\sigma}(\boldsymbol{\delta}) = \mathbf{s}$. When $\mathbf{s}$ approaches $\mathrm{bd}(\Delta_J)$, $|\nabla_{\mathbf{s}}\Omega(\mathbf{s})| = |\ln \mathbf{G}(\mathbf{s}) + \mathbf{1}|$ approaches infinity, which implies that $|\ln \mathbf{G}(\mathbf{s})|$ approaches infinity as well.

5.3 Proof of Proposition 2

The proof makes uses of Proposition 4 as applied to the function $\Omega : \mathbb{R}^{J+1} \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\Omega(\mathbf{s}) = \begin{cases} \sum_{j\in\mathcal{J}} s_j \ln G_j(\mathbf{s}) & \text{if } \mathbf{s} \in \Delta_J, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\ln \mathbf{G}$ satisfies Conditions *(GIL i)* – *(GIL iii)*.

I first show that $(\mathrm{int}(\Delta_J), \Omega)$ is a convex function of Legendre type. Ω is strictly convex on $\mathrm{int}(\Delta_J)$, since its Hessian is equal to $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}(\mathbf{s})$ for any $\mathbf{s} \in \mathrm{int}(\Delta_J)$ (see Lemma 4 in Fosgerau et al., 2019b). Ω is essentially smooth, since it is differentiable through the open convex set $\mathrm{int}(\Delta_J)$ with $\lim_{i\to\infty} |\nabla_{\mathbf{s}_i}\Omega(\mathbf{s}_i)| = +\infty$ whenever $\mathbf{s}_1, \mathbf{s}_2, \ldots$ is a sequence in $\mathrm{int}(\Delta_J)$ converging to a point $\mathbf{s} \in \mathrm{bd}(\Delta_J)$. The latter feature is shown by first noting that $\nabla_{\mathbf{s}}\Omega(\mathbf{s}) = \ln \mathbf{G}(\mathbf{s}) + \mathbf{1}$ for $\mathbf{s} \in \mathrm{int}(\Delta_J)$ and then using that $\lim_{\mathbf{s}\to \mathrm{bd}(\Delta_J)} |\ln \mathbf{G}(\mathbf{s})| = +\infty$.

Then, using Proposition 4, $\nabla_{\mathbf{s}}\Omega = \ln \mathbf{G} + \mathbf{1}$, and thus $\ln \mathbf{G}$, is a bijection between $\mathrm{int}(\Delta_J)$ and $\mathbb{R}^{J+1}$ with a continuous inverse mapping.

6 Conclusion

This paper studies the relationships between the class of GIL models developed by Fosgerau et al. (2019b) and the class of PUM studied by Allen and Rehbeck (2019a). It establishes that the conditions that the GIL models impose on the inverse demand function are necessary and sufficient for consistency with PUM maximization.

The GIL model can be used in different contexts. They can be used for demand estimation purposes and to study economic questions, such as mergers, products' entry, changes in regulations (e.g., taxes). In particular, Fosgerau et al. (2019b) build the IPDL model and show its use by estimating the demand for ready-to-eat cereals in Chicago in 1991-1992. Moreover, recalling that the inverse function is the target of estimation in structural demand estimation, one could use the shape restrictions imposed on GIL models for non-parametric estimation; alternatively, these conditions could be tested after estimation.

Lastly, the GIL model could also be used to model consumer's dynamic behavior in the spirit of De Groote and Verboven (2019) and matching in the spirit of Galichon and Salanié (2015).

Appendices

A Preliminaries

A.1 Elements of Convex Analysis

This subsection provides the elements of convex analysis used in the paper. See Rockafellar (1970) and Boyd and Vandenberghe (2004) for a comprehensive treatment of the topic.

Consider a convex function f that takes values in the extended real number line and whose domain of definition is a subset $\mathcal{X}$ of $\mathbb{R}^{J+1}$: $f : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$. Its *effective domain* $\mathrm{dom} f$ is defined by $\mathrm{dom} f = \{\mathbf{x} \in \mathcal{X} | f(\mathbf{x}) < +\infty\}$ and is a convex set in $\mathbb{R}^{J+1}$. The convexity of f is equivalent to that of the restriction of f to $\mathrm{dom} f$.

A *proper convex function* f is a convex function that takes values in the extended real number line such that $f(\mathbf{x}) < +\infty$ for at least one $\mathbf{x}$ and $f(\mathbf{x}) > -\infty$ for every $\mathbf{x}$. Then, f is proper if and only if its effective domain $\mathrm{dom} f$ is non-empty and the restriction of f to $\mathrm{dom} f$ is finite. In other words, a proper convex function on $\mathbb{R}^{J+1}$ is a function obtained by taking a finite convex function f on a non-empty convex set $\mathrm{dom} f$ and then extending it to all of $\mathbb{R}^{J+1}$ by setting $f(\mathbf{x}) = +\infty$ for $\mathbf{x} \notin \mathrm{dom} f$.

Let $f : \mathbb{R}^{J+1} \to \mathbb{R} \cup \{+\infty\}$. The *convex conjugate* of the function f is the function $f^* : \mathbb{R}^{J+1} \to \mathbb{R} \cup \{+\infty\}$ defined as

$$f^*(\mathbf{x}^*) = \sup_{\mathbf{x} \in \mathrm{dom} f} \{\mathbf{x}^{*\top}\mathbf{x} - f(\mathbf{x})\}.$$

Note that f^* is a convex function, regardless of whether f is convex. In addition, when f is convex, the subscript $\mathbf{x} \in \mathrm{dom} f$ is not necessary since, by definition, $\mathbf{x}^{*\top}\mathbf{x} - f(\mathbf{x}) = -\infty$ for $\mathbf{x} \notin \mathrm{dom} f$. The conjugate of a differentiable function f is also called the *Legendre transform* of f.

A proper convex function f is *essentially smooth* if (i) $\mathrm{int}(\mathrm{dom} f)$ is non-empty;

(ii) f is differentiable throughout $\mathrm{int}(\mathrm{dom} f)$, (iii) $\lim_{i\to\infty}|\nabla_{\mathbf{x}_i} f(\mathbf{x}_i)| = +\infty$ whenever $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is a sequence in $\mathrm{int}(\mathrm{dom} f)$ converging to a point $\mathbf{x} \in \mathrm{bd}(\mathrm{dom} f)$.

A pair $(\mathrm{int}(\mathrm{dom} f), f)$ is a *convex function of Legendre type* if $\mathrm{int}(\mathrm{dom} f)$ is an open convex set and f is a strictly convex function on $\mathrm{int}(\mathrm{dom} f)$ that is essentially smooth.

The results of this paper make use of the following two propositions.

Proposition 3 (Theorems 23.5 and 26.1 in Rockafellar (1970))**.** Let $f : \mathbb{R}^{J+1} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Assume that f is continuous and essentially smooth.The following five conditions are equivalent

1. $\mathbf{x}^* = \nabla f(\mathbf{x}), \quad \mathbf{x} \in \mathrm{int}(\mathrm{dom} f)$;
2. $\mathbf{x} = \nabla f^*(\mathbf{x}^*), \quad \mathbf{x}^* \in \mathrm{int}(\mathrm{dom} f^*)$;
3. $\mathbf{x} = \sup_{\mathbf{z}}\{\mathbf{x}^{*\top}\mathbf{z} - f(\mathbf{z})\}$;
4. $\mathbf{x}^* = \sup_{\mathbf{z}^*}\{\mathbf{z}^{*\top}\mathbf{x} - f^*(\mathbf{z}^*)\}$;
5. $f(\mathbf{x}) + f^*(\mathbf{x}^*) = \mathbf{x}^{*\top}\mathbf{x}$.

Proposition 4 (Theorem 26.5 in Rockafellar (1970))**.** Let $f : \mathbb{R}^{J+1} \to \mathbb{R} \cup \{+\infty\}$ be a continuous convex function. Assume that $(\mathrm{int}(\mathrm{dom} f), f)$ is a convex function of Legendre type. Then $(\mathrm{int}(\mathrm{dom} f^*), f^*)$ is also a convex function of Legendre type.

Furthermore, the gradient mapping $\nabla_{\mathbf{x}} f$ is a continuous bijection between $\mathrm{int}(\mathrm{dom} f)$ and $\mathrm{int}(\mathrm{dom} f^*)$, with a continuous inverse mapping $(\nabla_{\mathbf{x}} f)^{-1} = \nabla_{\mathbf{x}^*} f^*$, i.e., $(\nabla_{\mathbf{x}} f)^{-1}(\mathbf{x}^*) = \nabla_{\mathbf{x}^*} f^*(\mathbf{x}^*)$ for $\mathbf{x}^* \in \mathrm{int}(\mathrm{dom} f^*)$.

A.2 An Euler-Type Equation

The following lemma, presented in a slightly different version in Fosgerau et al. (2019b), establishes that the logarithm of a homogeneous function (which is a homothetic function) satisfies a modification of the generalized Euler equation for homothetic functions (McElroy, 1969).

Lemma 4. Consider a function $\mathbf{G} : [0,\infty)^{J+1} \to [0,\infty)^{J+1}$. Assume that $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ is symmetric on $\mathrm{int}(\Delta_J)$. $\mathbf{G}$ is homogeneous of degree K if and only if it satisfies the Euler-type equation

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} = K, \quad \mathbf{s} \in \mathrm{int}(\Delta_J). \tag{18}$$

Proof.
$\Rightarrow$ is shown by Fosgerau et al. (2019b) for the case of $K = 1$, but the proof can be easily extended to any K as follows

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_j(\mathbf{s})}{\partial s_k} = \sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_k(\mathbf{s})}{\partial s_j} = \frac{\sum_{j\in\mathcal{J}} s_j \frac{\partial G_k(\mathbf{s})}{\partial s_j}}{G_k(\mathbf{s})} = \frac{K G_k(\mathbf{s})}{G_k(\mathbf{s})} = K,$$

where the first equality uses the symmetry of $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ and the third equality uses uses the Euler equation for the homogeneous function $\mathbf{G}$.
$\Leftarrow$ Assume that $\mathbf{G}$ satisfies the Euler-type Equation (18). Then, by symmetry of $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial \ln G_k(\mathbf{s})}{\partial s_j} = K,$$

that is

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial G_k(\mathbf{s})}{\partial s_j} \frac{1}{G_k(\mathbf{s})} = K,$$

and

$$\sum_{j\in\mathcal{J}} s_j \frac{\partial G_k(\mathbf{s})}{\partial s_j} = K G_k(\mathbf{s}),$$

which implies that $\mathbf{G}$ is homogeneous of degree K. □

B On the Additive Random Utility Model (ARUM)

The ARUM was popularized by McFadden (1973). Due to its ability to model the behavior of heterogeneous consumers in the presence of many differentiated products in a tractable and parsimonious way, it has been ubiquitously used in the literature on structural estimation of demand models for differentiated products since Berry (1994) and Berry et al. (1995).

The ARUM In the ARUM, the utility u_j that each consumer derives from choosing product $j \in \mathcal{J}$ is the sum of a deterministic utility term δ_j and a random utility term ε_j

$$u_j = \delta_j + \varepsilon_j, \quad j \in \mathcal{J},$$

where δ_j is given by Equation (1).[19] This specification assumes, as it is the case for the PUM, that the only source of consumer heterogeneity in preferences is due to the vector of random utility terms $\boldsymbol{\varepsilon}$ whose distribution is parametrized by the vector $\boldsymbol{\theta}_2$.

As it is standard in the discrete choice literature, I further assume that the random vector $\boldsymbol{\varepsilon}$ follows a joint distribution with finite means that is absolutely continuous, independent of $\boldsymbol{\delta}$, and fully supported on $\mathbb{R}^{J+1}$. This implies that utility ties $u_i = u_j$, $i \neq j$, occur with probability 0, meaning that the argmax set of the ARUM is almost surely a singleton; that the choice probabilities are all almost everywhere positive; and that demands depend on prices and product characteristics only through product indexes $\boldsymbol{\delta}$.

In an ARUM, each consumer chooses the product that provides her the highest utility, so that demands are given by

$$\sigma_j(\boldsymbol{\delta}) = \Pr\left(u_j \geq u_i,\ \forall i \neq j\right), \quad j \in \mathcal{J}. \tag{19}$$

The Daly-Zachary Conditions Daly and Zachary (1979) provide the shape restrictions that any ARUM impose on their demands. These conditions, known as the Daly-Zachary conditions, are given as follows.

Conditions DZ The demand function $\boldsymbol{\sigma}$ satisfies, for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$ and for all $j \in \mathcal{J}$,

(DZ i) $\sigma_j(\boldsymbol{\delta}) > 0$ and $\sum_{k \in \mathcal{J}} \sigma_k(\boldsymbol{\delta}) = 1$ (positive unit demand)

(DZ ii) $\sigma_j(\boldsymbol{\delta}) = \sigma_j(\boldsymbol{\delta} + c\mathbf{1})$ for all $c \in \mathbb{R}$ (translation invariance)

(DZ iii) $\lim_{\delta_j \to \infty} \sigma_j(\boldsymbol{\delta}) = 1$ and $\lim_{\delta_j \to \infty} \sigma_i(\boldsymbol{\delta}) = 0$ with $\delta_i < \infty$, $i \neq j$ (boundary conditions)

(DZ iv) The partial derivatives of σ_j with respect to any set of distinct product indexes other than δ_j exist, are independent of the order of differentiation, and satisfy

$$(-1)^k \frac{\partial^k \sigma_j(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}_{-j}^k} > 0, \quad k = 1, \ldots, J, \tag{20}$$

with $\boldsymbol{\delta}_{-j}^k$ any k-subvector of $\boldsymbol{\delta}_{-j}$ (positivity)

[19] This is equivalent to let the residual income $(y - p_j)$ enter linearly with a coefficient α, which implies unit demand and absence of income effect.

(DZ v) $\mathbf{J}^{\delta}_{\sigma}$ is symmetric (Slutsky symmetry)

Note that, with respect to the original Daly-Zachary conditions, Conditions *(DZ i)* and *(DZ iv)* hold with strict inequality (i.e., the original Daly-Zachary conditions allow zero demands and zero partial derivatives). This is a consequence of the assumption that ε is fully supported on $\mathbb{R}^{J+1}$, which, e.g., is satisfied by the logit and nested logit models.

The following proposition, due to Daly and Zachary (1979) and restated by Anderson et al. (1992) in their Theorem 3.1, states that Conditions *(DZ i)* – *(DZ v)* are necessary and sufficient for consistency with ARUM maximization.[20]

Proposition 5. The ARUM choice probabilities (19) satisfy Conditions *(DZ i)* – *(DZ v)*. Conversely, any demand satisfying Conditions *(DZ i)* – *(DZ v)* can be derived as ARUM choice probabilities (19).

Proof. See Proofs of Theorem 3.1. in Anderson et al. (1992) or of Theorem 3 in Koning and Ridder (2003). □

Comparison with the MDZ Conditions As highlighted by Hofbauer and Sandholm (2002), ARUM satisfies Slutsky symmetry and positive definiteness, plus the additional positivity condition. This implies, as already noted by Koning and Ridder (2003), that Slutsky symmetry and positive definiteness is weaker than Slutsky symmetry and positivity.

Koning and Ridder (2003) show that Slutsky $\mathbf{J}^{\mathbf{p}}_{\sigma}$ symmetry and negative semi-definiteness is weaker than Slutsky symmetry and non-negativity (i.e., Equation (20) holds with weak inequality). As an illustration, they consider the simple $J+1=2$ case (see their Appendix C). Their illustration can be easily extended to show that Slutsky $\mathbf{J}^{\delta}_{\sigma}$ symmetry and positive definiteness is weaker than Slutsky symmetry and positivity.

[20]The Daly-Zachary conditions are also known as the Daly-Zachary-McFadden conditions, due to McFadden (1981). They were re-stated by Anderson et al. (1992) in their Theorem 3.1 and further studied by Koning and Ridder (2003). Other papers have investigated the restrictions on demands that make them consistent with a random utility maximization (see e.g., Armstrong and Vickers (2015), Jaffe and Weyl (2010), Jaffe and Kominers (2012), Bhattacharya (2019)).

Slutsky $\mathbf{J}_{\sigma}^{\delta}$ positive definiteness requires that the Slutsky matrix

$$\begin{pmatrix} \frac{\partial \sigma_0(\boldsymbol{\delta})}{\partial \delta_0} & \frac{\partial \sigma_0(\boldsymbol{\delta})}{\partial \delta_1} \\ \\ \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial \delta_0} & \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial \delta_1} \end{pmatrix}$$

be positive definite, which implies that the diagonal elements must be positive. Positivity requires that

$$\frac{\partial \sigma_0(\boldsymbol{\delta})}{\partial \delta_1} < 0 \quad \text{and} \quad \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial \delta_0} < 0.$$

Since Condition (i) holds, Slutsky symmetry implies

$$\frac{\partial \sigma_0(\boldsymbol{\delta})}{\partial \delta_0} = -\frac{\partial \sigma_0(\boldsymbol{\delta})}{\partial \delta_1} = -\frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial \delta_0} = \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial \delta_1}.$$

Assume that Condition (i) holds. Then, no-negativity and Slutsky $\mathbf{J}_{\sigma}^{\delta}$ symmetry imply that the Slutsky matrix is positive definite and symmetric, with the additional requirement that its off-diagonal elements negative. They are thus stronger than Slutsky $\mathbf{J}_{\sigma}^{\delta}$ symmetry and positive definiteness.

C The Logit Example

The standard logit model satisfies the properties of the ARUM defined in Appendix B. It rules out income effect and the only source of consumer heterogeneity in preferences is due to the vector of random utility terms $\boldsymbol{\epsilon}$ where ε_j are i.i.d. type I extreme value, so that $\boldsymbol{\epsilon}$ follows a joint distributes with finite means that is absolutely continuous, independent of $\boldsymbol{\delta}$, and fully supported on $\mathbb{R}^{J+1}$. This Appendix illustrates the findings of Section 4 through the logit example.

The Modified Daly-Zachary Conditions The logit model has demands given by

$$\sigma_j(\boldsymbol{\delta}) = \frac{e^{\delta_j}}{\sum_{k \in \mathcal{J}} e^{\delta_k}} = \frac{1}{1 + \sum_{k \in \mathcal{J} \setminus \{j\}} e^{\delta_k - \delta_j}}, \tag{21}$$

which satisfies the modified Daly-Zachary conditions. Indeed, demands (21) are positive and sum to one. They are invariant to translation in $\boldsymbol{\delta}$ since

$$\sigma_j(\boldsymbol{\delta}+c\mathbf{1}) = \frac{1}{1+\sum_{k\in\mathcal{J}\setminus\{j\}} e^{(\delta_k+c)-(\delta_j+c)}} = \sigma_j(\boldsymbol{\delta}),$$

and, with $\delta_k < \infty$, $k \neq j$, $\lim_{\delta_j\to\infty} \sigma_j(\boldsymbol{\delta}) = 1$ and $\lim_{\delta_j\to\infty} \sigma_k(\boldsymbol{\delta}) = 0$, $k \neq j$. Its Slutsky matrix given by

$$\mathbf{J}_{\sigma}^{\delta} = \begin{bmatrix} s_0(1-s_0) & -s_0 s_1 & \cdots & -s_0 s_J \\ -s_0 s_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -s_0 s_{J-1} \\ -s_0 s_J & \cdots & -s_0 s_{J-1} & s_J\left(1-s_J\right) \end{bmatrix},$$

with $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$, is positive definite on $\mathbb{R}_0^{J+1}$, since

$$\mathbf{z}^{\mathsf{T}}\mathbf{J}_{\sigma}^{\delta}\mathbf{z} = \sum_{i\in\mathcal{J}}\sum_{i<j} s_i s_j \left(z_i - z_j\right)^2 > 0, \quad \text{for all } \mathbf{z} \in \mathbb{R}_0^{J+1},$$

and symmetric.

Shape Restrictions on the Inverse Demand The logit model has inverse demands given by

$$\sigma_j^{-1}(\mathbf{s}) = \ln G_j(\mathbf{s}) + c, \quad j \in \mathcal{J},$$

with

$$\ln G_j(\mathbf{s}) \equiv \ln\left(s_j\right), \quad j \in \mathcal{J}.$$

where c is the consumer surplus, which is given by the log-sum-exp function

$$CS(\boldsymbol{\delta}) = \ln\left(\sum_{j\in\mathcal{J}} e^{\delta_j}\right),$$

up to an additive constant (see Berry, 1994). Then, $\ln \mathbf{G}(\mathbf{s}) \equiv \left(\ln(s_0), \ldots, \ln(s_J)\right)$ satisfies Conditions *(GIL i)* – *(GIL iii)*: $\mathbf{G}$ is continuously differentiable and homogenous of degree one on $\text{int}(\Delta_J)$, since for $\lambda > 0$, $\mathbf{G}(\lambda\mathbf{s}) = \mathbf{G}(\lambda\mathbf{s})$; $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}} = \mathbf{I}_{J+1}$ is positive definite and symmetric on $\text{int}(\Delta_J)$; and $|\ln \mathbf{G}(\mathbf{s})| = \sum_{j\in\mathcal{J}} |\ln(s_j)|$ approaches infinity as $\mathbf{s}$ approaches $\text{bd}(\Delta_J)$.

The Logit Model as a Perturbed Utility Model As shown by Anderson et al. (1988), the logit demands can also be obtained from a model of the form of Equation (6), where Ω is given by the Shannon entropy[21]

$$\Omega(\mathbf{s}) = \begin{cases} \sum_{j\in\mathcal{J}} s_j \ln(s_j) & \text{if } \mathbf{s} \in \Delta_J, \\ +\infty & \text{otherwise.} \end{cases}$$

More precisely, Anderson et al. (1988) consider a representative consumer choosing a vector of market shares, trading off variety against quality, so as to maximize her utility composed of two terms: the first term, $\sum_{j\in\mathcal{J}} \delta_j s_j$, captures the net utility that she derives from consuming $\mathbf{s}$ in absence of interaction among products and the second term, $-\Omega(\mathbf{s})$, expresses her taste for variety. Indeed, if her utility were only composed of the first term, then she would only choose the product j with the highest net utility δ_j. Likewise, if her utility were only composed of the second term, then she would choose all the products with equal shares (and she would minimize her utility by choosing only one product). This last feature justifies why $-\Omega(\mathbf{s})$ expresses taste for variety and is a manifestation of the IIA property of the logit model.

The logit model satisfies Conditions *(PUM i)* – *(PUM iii)*. Indeed, the Shannon entropy Ω is continuously differentiable and strictly convex since its Hessian, equal to I_{J+1}, is positive definite; $|\nabla_{\mathbf{s}}\Omega(\mathbf{s})| = \sum_{j\in\mathcal{J}} |\ln(s_j)+1|$ approaches infinity as $\mathbf{s} \in \mathrm{bd}(\Delta_J)$.

[21] Note that it is known, from the convex analysis literature, that CS is the convex conjugate of the Shannon entropy (see Example 3.25 (p. 93) of Boyd and Vandenberghe, 2004).

Bibliography

Agranov, M. and P. Ortoleva (2017): "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 125, 40–68.

Allen, R. and J. Rehbeck (2018): "Hicksian Complementarity and Perturbed Utility Models," *Available at SSRN 2801345*.

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

——— (2019b): "Revealed Stochastic Choice with Attributes," *Available at SSRN 2818041*.

Anderson, S. P., A. de Palma, and J. F. Thisse (1988): "A representative consumer theory of the logit model," *International Economic Review*, 461–466.

——— (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.

Berry, S., A. Gandhi, and P. Haile (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

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

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

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 25, 242–262.

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

Bhattacharya, D. (2019): "Income Effects and Rationalizability in Multinomial Choice Models," *arXiv preprint arXiv:1902.11017*.

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

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019): "Deliberately stochastic," *American Economic Review*, 109, 2425–45.

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7, 83–115.

Clark, S. A. (1990): "A concept of stochastic transitivity for the random utility model," *Journal of Mathematical Psychology*, 34, 95–108.

Compiani, G. (2019): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," *Unpublished*.

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.

De Groote, O. and F. Verboven (2019): "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems," *American Economic Review*, 109, 2137–72.

Fosgerau, M., E. Melo, A. De Palma, and M. Shum (2019a): "Discrete Choice and Rational Inattention: A General Equivalence Result," *Available at SSRN: https://ssrn.com/abstract=2889048*.

Fosgerau, M., J. Monardo, and A. de Palma (2019b): "The Inverse Product Differentiation Logit Model," *Available at SSRN: https://ssrn.com/abstract=3141041*.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

Galichon, A. and B. Salanié (2015): "Cupids invisible hand: Social surplus and identification in matching models," *Available at SSRN 1804623*.

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

Hurwicz, L. (1971): "On the integrability of demand functions," *Preferences, utility and demand*.

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

Jaffe, S. and E. G. Weyl (2010): "Linear demand systems are inconsistent with discrete choice," *The BE Journal of Theoretical Economics*, 10.

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

Lewbel, A. (2001): "Demand Systems with and without Errors," *American Economic Review*, 91, 611–618.

Machina, M. J. (1985): "Stochastic choice functions generated from deterministic preferences over lotteries," *The economic journal*, 95, 575–594.

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. (1973): "Conditional logit analysis of qualitative choice behavior,"
.

——— (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.

McFadden, D. L. and M. Fosgerau (2012): "A theory of the perturbed consumer with general budgets," National Bureau of Economic Research.

Monardo, J. (2019): "The Flexible Inverse Logit Model," *Unpublished*.

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

RIESKAMP, J., J. R. BUSEMEYER, AND B. A. MELLERS (2006): "Extending the bounds of rationality: Evidence and theories of preferential choice," *Journal of Economic Literature*, 44, 631–661.

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.

VERBOVEN, F. (1996): "International price discrimination in the European car market," *The RAND Journal of Economics*, 240–268.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences de l'homme et de la société (SHS)

Titre : Essais sur l'estimation structurelle de la demande.

Mots clés : Econométrie structurelle ; économétrie industrielle empirique ; estimation de la demande.

Résumé :
L'estimation structurelle des modèles de demande sur des marchés de produits différenciés joue un rôle important en économie. Elle permet de mieux comprendre les choix des consommateurs et, entre autres, de mesurer les effets d'une fusion d'entreprise, de l'introduction d'un nouveau produit sur le marché ou d'une nouvelle régulation. L'approche traditionnelle consiste à spécifier un modèle d'utilité, typiquement un modèle d'utilité aléatoire additif, à en calculer ses demandes et à inverser ces dernières pour obtenir des équations de demande inverse qui serviront de base pour l'estimation. Toutefois, en général, ces demandes inverses n'ont pas de forme analytique. L'estimation exige donc une inversion numérique et l'emploi de procédures d'estimation non-linéaire, qui peuvent être difficiles à mettre en oeuvre et chronophages.
Cette thèse adopte une approche différente, en développant de nouveaux modèles de demande inverse qui sont cohérents avec un modèle d'utilité de consommateurs hétérogènes. Cette approche permet de capter de façon plus flexible les substitutions entre les produits, grâce à de simples régressions linéaires basées sur des données incluant les parts de marché, les prix et les caractéristiques des produits.
Le premier chapitre de cette thèse développe le modèle *inverse product differentiation logit* (IPDL), qui généralise les modèles logit emboîtés, permettant ainsi de capter de façon flexible les substitutions entre les produits, y compris de la complémentarité. Il montre que le modèle IPDL appartient à une classe de modèles de demande inverse, nommée *generalized inverse logit* (GIL), laquelle inclut une grande majorité de modèles d'utilité aléatoire additifs qui ont été utilisés à des fins d'estimation de la demande.
Le second chapitre développe le modèle *flexible inverse logit* (FIL), un modèle GIL qui utilise une structure de nids flexible avec un nid pour chaque pair de produits. Il montre que le modèle FIL, projeté dans l'espace des caractéristiques des produits, permet d'obtenir des élasticités-prix qui dépendent directement des caractéristiques des produits et, en utilisant des simulations de Monte-Carlo, qu'il est capable de reproduire celles du "flexible" modèle logit à coefficients aléatoires.
Le troisième chapitre étudie la micro-fondation du modèle GIL. Il montre que les restrictions que le modèle GIL impose sur la fonction de demande inverse sont des conditions nécessaires et suffisantes de cohérence avec un modèle de consommateurs hétérogènes maximisant leur fonction d'utilité, connu sous le nom de *perturbed utility model* (PUM). Il montre également que tout modèle GIL génère une fonction de demande qui satisfait une légère variante des conditions de Daly-Zachary, laquelle permet de combiner substituabilité et complémentarité en demande.

Title : Essays on Structural Demand Estimation.

Keywords : Structural econometrics; empirical industrial organization; demand estimation.

Abstract : Estimation of structural demand models in differentiated product markets plays an important role in economics. It allows to better understand consumers' choices and, amongst other, to assess the effects of mergers, new products, and changes in regulation. The standard approach consists in specifying a utility model, typically an additive random utility model, computing its demands, and inverting them to obtain inverse demand equations, which will serve as a basis for estimation. However, since these inverse demands have generally no closed form, estimation requires numerical inversion and non-linear optimization, which can be painful and time-consuming.
This dissertation adopts a different approach, developing novel inverse demand models, which are consistent with a utility model of heterogeneous consumers. This approach allows to accommodate rich substitution patterns thanks to simple linear regressions with data on market shares, prices and product characteristics.
The first chapter of this dissertation develops the inverse product differentiation logit (IPDL) model, which generalizes the nested logit models to allow for richer substitution patterns, including complementarity. It also shows that the IPDL model belongs to the class of generalized inverse logit (GIL) models, which includes a vast majority of additive random utility models that have been used for demand estimation purposes.
The second chapter develops the flexible inverse logit (FIL) model, a GIL model that uses a flexible nesting structure with a nest for each pair of products. It shows that the FIL model, projected into product characteristics space, makes the price elasticities depending on product characteristics directly and, using Monte Carlo simulations, that it is able to mimic those from the "flexible" random coefficient logit model.
The third chapter studies the micro-foundation of the GIL model. It shows that the restrictions that the GIL model imposes on the inverse demand function are necessary and sufficient for consistency with a model of heterogeneous and utility-maximizing consumers, called perturbed utility model. It also shows that any GIL model yields a demand function that satisfies a slight variant of the Daly-Zachary conditions, which allows to combine substitutability and complementarity in demand.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France