THEMA

théorie économique,
modélisation et applications

The Inverse Product Differentiation Logit Model

Mogens Fosgerau, Julien Monardo, André de Palma

CY UNIVERSITÉ

cnrs

The Inverse Product Differentiation Logit Model[*]

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

November 29, 2022

Abstract

We introduce the inverse product differentiation logit (IPDL) model, a micro-founded inverse market share model for differentiated products that captures market segmentation according to one or more characteristics. The IPDL model generalizes the nested logit model to allow richer substitution patterns, including complementarity in demand, and can be estimated by linear instrumental variables regression using market-level data. Furthermore, we provide Monte Carlo experiments that compare the IPDL model to the workhorse empirical models of the literature. Lastly, we show the empirical

performance of the IPDL model using a well-known dataset on the ready-to-eat cereals market.

(JEL: C26, D11, D12, L)

Keywords: Demand estimation, Inverse demand, Logit, Consumer model, Differentiated products

1 Introduction

We introduce the inverse product differentiation logit (IPDL) model, a micro-founded inverse market share model that captures market segmentation according to one or more characteristics. The IPDL model generalizes the nested logit model to allow richer substitution patterns while retaining its attractive features: it can easily be estimated by linear instrumental variables regression using market-level data; and it is consistent with a model of heterogeneous, utility-maximizing consumers, which makes it useful for analyzing a wide range of economic questions.

The nested logit model is commonly used to estimate demand in differentiated products markets. It can be estimated by linear instrumental variable regression, using that its inverse market share function is linear-in-parameters and in closed form (Berry, 1994). However, it only captures market segmentation according to one or several characteristics treated hierarchically (i.e., partitioning the choice set into nests, nests into subsets, etc.), which imposes strong restrictions on substitution patterns. To avoid these restrictions, several Generalized Extreme Value (GEV) models (McFadden, 1978) have been proposed, for which, however, a closed-form inverse market share function does not exist, preventing the use of regression techniques.[1] Notably, Bresnahan et al. (1997)'s Product Differentiation Logit (PDL) model generalizes the nested logit model by treating the grouping characteristics non-hierarchically. The IPDL model allows the same grouping structure as the PDL model but builds it into a linear-in-parameters, closed-form inverse market share function rather than into a closed-form market share function. The IPDL model is thus not the inversion of the PDL model or any other GEV model but a novel model we introduce in this paper.

---

[1]See Subsection 3.3 for further details on GEV models.

2

The state-of-the-art approach to estimating demand in differentiated products markets is the random coefficient logit (RCL) model with structural error terms to allow for unobserved product characteristics, estimated using the methodology developed by Berry et al. (1995, hereafter BLP).[2] The RCL model allows for rich substitution patterns determined by a random coefficient specification of the distribution of unobserved preference heterogeneity. The methodology developed by BLP involves a non-linear, non-convex optimization problem and the simulation and numerical inversion of the market share function. With the IPDL model, in contrast, we specify directly a closed-form, linear-in-parameters inverse market share function, which generates substitution patterns determined by segmentation of the differentiated products, and which can be estimated by linear instrumental variables regression.

By specifying an inverse market share function, we avoid some restrictions embedded in the GEV and the RCL models. In particular, these latter models restrict products to be substitutes in demand (i.e., a positive cross-price derivative of market share). By contrast, the IPDL model allows for complementarity in demand. Furthermore, we show that the IPDL model is consistent with utility maximization, which makes it useful for performing counterfactual analyses such as merger simulation. We show the IPDL model is consistent with a representative consumer who chooses a vector of market shares to maximize her quasi-linear direct utility function subject to a budget constraint, which is, in turn, consistent with a population of utility-maximizing, heterogeneous consumers.

However, relying on an inverse market share function entails a cost. Without additional assumptions about the distribution of preferences in the population of consumers, the IPDL model cannot be used to address economic questions at the individual level, such as the distributional effects of any events or policies. Fortunately, many economic questions of interest do not require knowing the distribution

---

[2]BLP provide an estimator that allows for rich substitution patterns while handling the endogeneity issues related to the modelling of unobserved product characteristics. BLP also propose an algorithm to compute that estimator. Conlon and Gortmaker (2020) consolidate best estimation practices in a Python package. Dubé et al. (2012) propose another algorithm to compute the BLP estimator. Lee and Seo (2015) and Salanié and Wolak (2022) provide approximations of the BLP estimator that are faster and easier to compute.

of preferences and can thus be addressed using the IPDL model. Prominent examples include the measurement of market power, the welfare effects of a merger, a new product introduced to the market, or regulatory changes, such as tax or trade policies.

We investigate the empirical properties of the IPDL model using Monte Carlo experiments. The IPDL model performs well in approximating the substitution patterns generated by the PDL model and the RCL model with independent normal random coefficients on dummies for groups. Furthermore, even without complementarity in demand, the IPDL model allows substitution patterns that the RCL model cannot replicate. We also find that the IPDL model outperforms the RCL model in approximating the substitution patterns generated by the PDL model. While there is a concern that the IPDL model may generate complementarity when there is none in the data, we do not observe this in our experiments.

We then analyze the empirical performance of the IPDL model using a well-known dataset on the ready-to-eat cereals market, which exhibits segmentation according to the brand name of the cereals and the market segment they belong to. We estimate the corresponding IPDL model and compare to the RCL model with independent normal random coefficients on dummies for groups. We estimate two specifications of both models: one with many markets and another with many products. In both specifications, the IPDL model provides a better fit to the data than the RCL model. Moreover, in both specifications, the IPDL model generates significantly higher substitution between cereals than the RCL model. The IPDL model implies markups in line with the literature (Nevo, 2001; Michel et al., 2022), while the RCL model, in the specification with many products, implies significantly lower markups.

The Monte Carlo experiments and the empirical application suggest that the IPDL model is particularly useful in describing markets that exhibit segmentation. This is the case of many markets: the beers market is segmented by brand and style (e.g., IPA, Lager, etc.), the cars market by brand and market segment (e.g., compact, luxury), the ready-to-eat cereals market by brand and market segment (e.g., kids, adults). Often market segmentation proxies for other (continuous) product characteristics, including prices: lager beers have lower alcohol content than IPA

4

beers; luxury cars are less fuel efficient, larger, and more expensive than compact cars; cereals for kids are more sugary than cereals for adults. In these cases, the IPDL model can generate substitution patterns determined, at least indirectly, by these continuous characteristics. When there is no obvious market segmentation, it is still possible to use a clustering algorithm to define groups for the IPDL model.

Other approaches exist for estimating demand in differentiated products markets based on market-level data (e.g., Barnett and Serletis, 2008; Nevo, 2011; Gandhi and Nevo, 2021). The flexible functional form approach (e.g., the AIDS model of Deaton and Muellbauer, 1980) provides rich substitution patterns, including complementarity in demand, and has been successfully applied to many economic questions. However, in this approach, the unobservables enter in a very restrictive way, there are many parameters to estimate, and the introduction of new products cannot be addressed. Other authors propose semi- or non-parametric demand models (e.g., Pinkse and Slade, 2004; Haag et al., 2009; Blundell et al., 2012; Compiani, 2022) or the use a more flexible specification of the distribution of the random coefficients in the RCL model (e.g., Lu et al., 2022; Wang, 2022). Closest to our paper is Compiani (2022), who non-parametrically estimates inverse market share functions for differentiated products based on market-level data. His approach provides rich substitution patterns but faces the curse of dimensionality that restricts its use to small choice sets. By contrast, the IPDL model can handle very large choice sets.

The paper is organized as follows. Section 2 presents our general setting and discusses the role of demand inversion. Section 3 introduces the IPDL model, studies its properties, and discusses estimation and identification with market-level data. It further provides Monte Carlo experiments that compare the IPDL model to the workhorse empirical models of the literature and an extended discussion of its relationship to other demand models. Section 4 analyzes the empirical performance of the IPDL model using a well-known dataset on the ready-to-eat cereals markets and compare to the RCL model estimated using the methodology developed by BLP. Section 5 concludes.

2 General Setting

We first introduce our setting and discuss the role of demand inversion in estimation. Consider a population of consumers making choices among a set of $J+1$ differentiated products, indexed by $\mathcal{J} = \{0, 1, \ldots, J\}$, where product $j = 0$ is the outside good. We consider data on market shares s_{jt}, prices p_{jt}, and K product/market characteristics $\mathbf{x}_{jt}$ for each product $j = 1, \ldots, J$ in each market $t = 1, \ldots, T$ (Berry, 1994; Berry et al., 1995; Nevo, 2001; Berry and Haile, 2021). For each market t, the market shares s_{jt} are positive and sum to 1, i.e., $\mathbf{s}_t = (s_{0t}, \ldots, s_{Jt}) \in \Delta^\circ$, where Δ° is the set of positive market share vectors.[3]

Following 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 an unobserved characteristics term for product/market jt, and where $\boldsymbol{\theta}_1$ is a vector of parameters. Consider the system of market share equations

$$s_{jt} = \sigma_j\left(\boldsymbol{\delta}_t; \boldsymbol{\theta}_2\right), \quad j \in \mathcal{J}, \quad t = 1, \ldots, T, \tag{1}$$

which relates the vector of observed market shares, $\mathbf{s}_t$, to the vector of product indexes $\boldsymbol{\delta}_t = (\delta_{0t}, \ldots, \delta_{Jt})^\intercal$, through the market share function $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_J)$, where $\boldsymbol{\theta}_2$ is a vector of parameters.

Normalize the index of the outside good by setting $\delta_{0t} = 0$ in each market t so that $\boldsymbol{\delta}_t \in \mathcal{D} \equiv \{\boldsymbol{\delta}_t \in \mathbb{R}^{J+1} : \delta_{0t} = 0\}$, and assume that the function $\boldsymbol{\sigma}(\cdot; \boldsymbol{\theta}_2) : \mathcal{D} \to \Delta^\circ$ is invertible. Then, the inverse market share function, denoted by σ_j^{-1}, maps from market shares $\mathbf{s}_t$ to each index δ_{jt} with

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

---

[3]Formally, Δ° is the relative interior of the unit simplex in $\mathbb{R}^{J+1}$ denoted by $\Delta = \left\{ \mathbf{s} \in [0, \infty)^{J+1} : \sum_{j \in \mathcal{J}} s_j = 1 \right\}$.

In addition, assume a linear index,

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

where the vector of parameters $\boldsymbol{\beta} \in \mathbb{R}^K$ captures the consumers' taste for characteristics $\mathbf{x}_{jt}$ and the parameter $\alpha > 0$ is the consumers' marginal utility of income. Then the unobserved product characteristics terms, ξ_{jt}, can be written as a 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 product characteristics terms, ξ_{jt}, are the structural error terms of the model, as they are observed by consumers and firms but not by the modeler. 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 the market share function of each product depends on the entire vector of endogenous prices and unobserved product characteristics. Then, following Berry (1994), we can estimate the market share function $\boldsymbol{\sigma}$ based on the conditional moment restrictions $\mathbb{E}[\xi_{jt}|\mathbf{z}_t] = 0$ for all $j \in \mathcal{J}$ and $t = 1, \ldots, J$, provided that there exist appropriate instruments $\mathbf{z}_t$ for prices and market shares.

Since the seminal papers by Berry (1994) and Berry et al. (1995), the standard practice of the demand estimation literature with market-level data has been to specify a GEV or RCL model. For these models, except for the logit and nested logit models, the implied inverse market share function is not in closed form and must then computed numerically during estimation, which prevents the use of standard regression techniques. In this paper, we instead directly specify a closed-form, invertible, and linear-in-parameters inverse market share function, for which estimation amounts just to linear regression.

Consider as an example the three-level nested logit model, which partitions the choice set into nests and nests into subnets. This is a special case of the IPDL model that we introduce in this paper. Let $\boldsymbol{\theta}_2 = (\mu_1, \mu_2)$ be the vector of grouping parameters, with $\sum_{d=1}^2 \mu_d < 1$, $\mu_1 \geq 0$ and $\mu_2 \geq 0$ to make the nested logit model

consistent with utility maximization. The corresponding inverse market share function is linear in parameters:

$$\sigma_j^{-1}\left(\mathbf{s}_t; \boldsymbol{\theta}_2\right) = \left(1 - \sum_{d=1}^{2} \mu_d\right) \ln\left(s_{jt}\right) + \sum_{d=1}^{2} \mu_d \ln\left(s_{d(j),t}\right) + c_t = \delta_{jt}, \quad (4)$$

where $s_{1(j),t} = \sum_{k \in 1(j)} s_{kt}$ and $s_{2(j),t} = \sum_{k \in 2(j)} s_{kt}$, with $1(j)$ and $2(j)$ the sets of products belonging the same nest and to the same subnest as product j, respectively,[4] and where $c_t \in \mathbb{R}$ is a market-specific constant determined by the normalization of the vector $\boldsymbol{\delta}_t$. The three-level nested logit model corresponds to the logit model when $\mu_1 = 0$ and $\mu_2 = 0$ and to the two-level nested logit model when $\mu_1 = 0$ or $\mu_2 = 0$.

Assume that the outside good is in a nest by itself, such that $\sigma_0^{-1}(\mathbf{s}_t; \mu_1, \mu_2) = \ln(s_{0t}) + c_t = \delta_{0t}$, and, in turn, as for the logit model, $c_t = -\ln(s_{0t})$ since $\delta_{0t} = 0$. Then, combining with Equation (4), the three-level nested logit model boils down to the linear regression model (Verboven, 1996a)

$$\ln\left(\frac{s_{jt}}{s_{0t}}\right) = \mathbf{x}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \sum_{d=1}^{2} \mu_d \ln\left(\frac{s_{jt}}{s_{d(j),t}}\right) + \xi_{jt}, \quad (5)$$

for all products $j = 1, \ldots, J$ in each market $t = 1, \ldots, T$, which requires at least one instrument for price and two for the endogenous log-share terms for identification.

3   The IPDL Model

The nested logit model can be estimated by linear instrumental variable regression and, due to its parsimony, can handle very large choice sets. However, it imposes strong restrictions on substitution patterns. In this section, we introduce the IPDL model, which generalizes the inverse market share function of the nested logit model while maintaining its desirable features.

---

[4]Setting $\sigma_1 = \mu_1 + \mu_2$ and $\sigma_2 = \mu_1$, where σ_1 and σ_2 refer to as grouping parameters for subnests and nests respectively, we recover Equation (10) of Verboven (1996a) with $0 \leq \sigma_2 \leq \sigma_1 < 1$.

Setting Suppose that each market exhibits product segmentation according to D discrete product characteristics, indexed by d. Each of these *grouping characteristics* d defines a partition of the set of products, that is, a finite number of *groups* of products such that each product belongs to exactly one group for each grouping characteristic. For example, cars may be grouped by brand, size, and fuel type. We denote by $d(j) \subseteq \{1, \ldots, J\}$ the set of products grouped with product j according to grouping characteristic d. The grouping structure is assumed to be exogenous and common across markets.

Let $\boldsymbol{\theta}_2 = (\mu_1, \ldots, \mu_D)$ be the vector of grouping parameters, with $\sum_{d=1}^{D} \mu_d < 1$ and $\mu_d \geq 0$, $d = 1, \ldots, D$ to make the IPDL model consistent with utility maximization, as we show below. The IPDL model has an inverse market share function defined by

$$\sigma_j^{-1}(\mathbf{s}_t; \boldsymbol{\theta}_2) = \left(1 - \sum_{d=1}^{D} \mu_d\right) \ln(s_{jt}) + \sum_{d=1}^{D} \mu_d \ln\left(s_{d(j),t}\right) + c_t = \delta_{jt}, \quad j = 1, \ldots, J,$$

(6)

where we recall that $s_{d(j),t} = \sum_{k \in d(j)} s_{kt}$ is the market share of the group $d(j)$.

Two products are of the same *type* if they belong to the same group according to all grouping characteristics d. We assume that the outside good is the only product of its type, that is,

$$\sigma_0^{-1}(\mathbf{s}_t; \boldsymbol{\theta}_2) = \ln(s_{0t}) + c_t = \delta_{0t}.$$

(7)

The index for the outside good is normalized to zero, $\delta_{0t} = 0$, and we find that $c_t = -\ln(s_{0t})$.

By construction, the logit and the nested logit models are special cases of the IPDL model: the logit model is obtained when there is no product segmentation, and the nested logit model is obtained when the grouping structure is hierarchical.

The IPDL model generalizes the inverse market share function of the nested logit model by allowing arbitrary, non-hierarchical grouping structures, that is, any partition of the set of products for each grouping characteristic. In Subsection 3.1, we show that the non-hierarchical grouping structure allows the IPDL model to accommodate richer substitution patterns than the nested logit model. The product differentiation logit (PDL) model of Bresnahan et al. (1997) allows the same

non-hierarchical grouping structure but is a specific member of the family of GEV models. The IPDL model is different: in general, its inverse market share function does not correspond to any other model. In Subsection 3.1, we show the IPDL model avoids the restriction, inherent in the GEV and RCL models, that all products are substitutes in demand.

Note that the inverse market share function $\boldsymbol{\sigma}^{-1} = (\sigma_0^{-1}, \ldots \sigma_J^{-1})$ with elements given by Equations (6) and (7) is invertible.[5] That is, any vector of observed market shares $\mathbf{s}_t \in \Delta^\circ$ is rationalized by a unique vector of product indexes $\boldsymbol{\delta}_t \in \mathcal{D}$, which is key for identification purposes. However, the market share function of the IPDL model is not in closed form. Counterfactual analyzes typically require computing the market share function. This can be done by inverting the inverse market share function or by solving the utility maximization program (see below), numerically after estimation.

Micro-foundation   In Appendix B.2, we show that the IPDL model is consistent with a representative consumer model with taste for variety, such as the logit and nested logit models (Anderson et al., 1988; Verboven, 1996b). Specifically, the IPDL model is consistent with a representative consumer, endowed with income y, who chooses a vector $\mathbf{s}_t \in \Delta^\circ$ of positive market shares in market t so as to maximize her utility function given by

$$u(\mathbf{s}_t) \equiv \alpha y + \sum_{j \in \mathcal{J}} \delta_{jt} s_{jt} - \mu_0 \sum_{j \in \mathcal{J}} s_{jt} \ln(s_{jt}) - \sum_{d=1}^{D} \mu_d \left(\sum_{g \in d \cup \{0\}} s_{gt} \ln(s_{gt}) \right), \quad (8)$$

where $\mu_0 \equiv 1 - \sum_{d=1}^{D} \mu_d$, $s_{gt} \equiv \sum_{k \in g} s_{kt}$, and d is identified with the set of groups for grouping characteristic d. The second term in Equation (8) captures the net utility derived from the consumption of $\mathbf{s}_t$ absent interaction among products, and the remaining terms express taste for variety. Specifically, the parameter μ_0 measures taste for variety over the entire choice set, while each parameter μ_d measures taste for variety across groups according to characteristic d. A higher value of μ_d puts more weight on variety at the group level, which can be interpreted as meaning that

---

[5]See Lemma 3 in Appendix B.1.

products in the same group according to d are more similar. See Verboven (1996b) for a similar interpretation of the grouping parameter in the nested logit model.

Furthermore, we can show that the utility function (8) belongs to the class of utilities studied by Allen and Rehbeck (2019b), which can be interpreted as representing the behavior of heterogeneous, utility-maximizing consumers. With this interpretation, the grouping parameters capture consumer heterogeneity in taste, like the grouping parameters in the random utility interpretation of the nested logit model.

The IPDL model avoids restrictions inherent in the GEV and RCL models. In particular, these latter models assume that each consumer chooses the product that provides her with the highest utility among all the available products. This assumption, known as the single-unit purchase assumption, restricts products to be substitutes in demand (i.e., a positive cross-price derivative of market share). In contrast, we do not retain the single-unit purchase assumption. As a result, the IPDL model allows for complementarity in demand (see Subsection 3.1) and does not rule out multiple choices by individuals. Some instances of the IPDL model may be consistent with the single-unit purchase assumption, e.g., when the IPDL model is equivalent to the logit or nested logit models.

Identification and Estimation Combining Equations (6) and (7) and using that $\delta_{0t} = 0$ for all $t = 1, \ldots, T$, the IPDL model boils down to the linear regression of market shares on product characteristics, prices, and log-share terms

$$\ln \left(\frac{s_{jt}}{s_{0t}} \right) = \mathbf{x}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \sum_{d=1}^{D} \mu_d \ln \left(\frac{s_{jt}}{s_{d(j),t}} \right) + \xi_{jt}, \qquad (9)$$

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

Equation (9) has the same form as the logit and nested logit equations, except for the log-share terms. Following the literature, we assume that product characteristics $\mathbf{x}_{jt}$ are exogenous and that prices and log-share terms are endogenous. As a consequence, the IPDL model reduces to a linear instrumental variable regression, where identification requires at least one instrument for price and one for each of the log-share terms.

As it is well known, instruments for prices include cost shifters and markup shifters (see, e.g., Berry and Haile, 2014, 2016). The first set of instruments includes the Hausman instruments, i.e., prices in other markets (Hausman et al., 1994; Nevo, 2001). The second set of instruments includes the BLP instruments, i.e., functions of the characteristics of competing products (Berry et al., 1995; Gandhi and Houde, 2021) and exogenous market shocks such as mergers (Miller and Weinberg, 2017). Following Verboven (1996a) and Bresnahan et al. (1997), the BLP instruments for the IPDL model include, for each grouping characteristic, the sums of characteristics of other products belonging to the same group, the sums of characteristics of other products belonging to different groups or, alternatively, the corresponding squared differences in those characteristics.

Identification of grouping parameters μ_d requires exogenous variation in the relative share $s_{jt}/s_{d(j)}$. Intuitively, since they drive substitution patterns among products, their identification requires instruments that provide exogenous variation in the choice set, including changes in prices. Thus, both cost shifters and markup shifters are good candidates for instrumenting the log-share terms.

3.1 Substitution Patterns

The richness of the substitution patterns allowed by the IPDL model can be assessed by analyzing the matrix of own- and cross-price elasticities of demand as well as the matrix of diversion ratios. We derive these in Appendix B.3.

We first focus on the price elasticities of market shares. The cross-price elasticity from product j to product k is the percentage change in the market share of product k following a one-percent increase in the price of product j.

To better understand substitution in the IPDL model, consider the cereals market segmented by brand (General Mills, Kellogg's, Quaker) and market segment (all-family, adults, kids). Recall that two products are of the same type if they belong to the same group based on each characteristic, which here means they are of the same brand and market segment. Consider a price decrease of a cereal for kids sold by Kellogg's. We show that this price decrease will reduce the market share of all other cereals of a given type by the same percentage. In other words, the price

decrease will draw proportionately from all the cereals of a given type. It will, for example, reduce the market share of each cereal for adults sold by General Mills by the same percentage or of each cereal for all-family sold by Kellogg's by (in general another) constant percentage. This substitution pattern is a manifestation of the Independence from Irrelevant Alternatives (IIA) property among products of the same type. Furthermore, this price decrease will draw proportionately more or less from cereals of different types depending on the value of the grouping parameters and the number of groups these cereals share. This means that the IIA property does not hold in general for products of different types. There will be as many different cross-price elasticities per product as there are different product types.

Turn now to diversion ratios, which offer a better description of substitution patterns than cross-price elasticities (Conlon and Mortimer, 2021).[6] The diversion ratio from product j to product k is the fraction of consumers leaving product j following a price increase of product j, who switch to product k.

We use simulations to investigate the patterns of diversion ratios in the IPDL model. We randomly generate $1,000$ markets with $J = 45$ products exhibiting product segmentation according to two grouping characteristics with corresponding grouping parameters μ_1 and μ_2, each forming two groups so that there are four product types. Figure 1 summarizes the results. For clarity, we group products according to whether they belong to the same groups according to both grouping characteristics, belong to the same group only according to the first grouping characteristic, belong to the same group only according to the second grouping characteristic, or do not belong to the same group according to either grouping characteristics.

Figure 1 exhibits some clear and intuitive patterns. The diversion ratio is highest between products of the same type. It is second highest between products of different types but belonging to the same group according to the grouping characteristic with the largest grouping parameter. The diversion ratio is lowest for products of completely different types, i.e., that do not belong to the same group according to

---

[6]Consider Conlon and Mortimer (2021)'s example with three products: the first has a cross-price elasticity with the third of 0.034 and a market share of 0.1, whereas the second has a cross-price elasticity with the third of 0.01 and a market share of 0.35. More consumers switch to the second product than to the first as the price of the third product increases, even though the first product has a larger cross-price elasticity.

any grouping characteristic. Moreover, the diversion ratio between products of the same type increases with the grouping parameters. Conversely, for products of completely different types, the diversion ratio decreases with the grouping parameters and becomes negative when the grouping parameters are sufficiently large, making these products complements in demand.

In the supplement, we provide simulation results investigating the substitution patterns of the IPDL model. We find that products of the same type are always substitutes, while products of different types may be substitutes or complements, and that products closer in the characteristics space used to form product types (i.e. higher values of the grouping parameters and/or whether products belong to the same groups or not) have higher diversion ratios.

Figure 1: Diversion Ratios in the IPDL Model
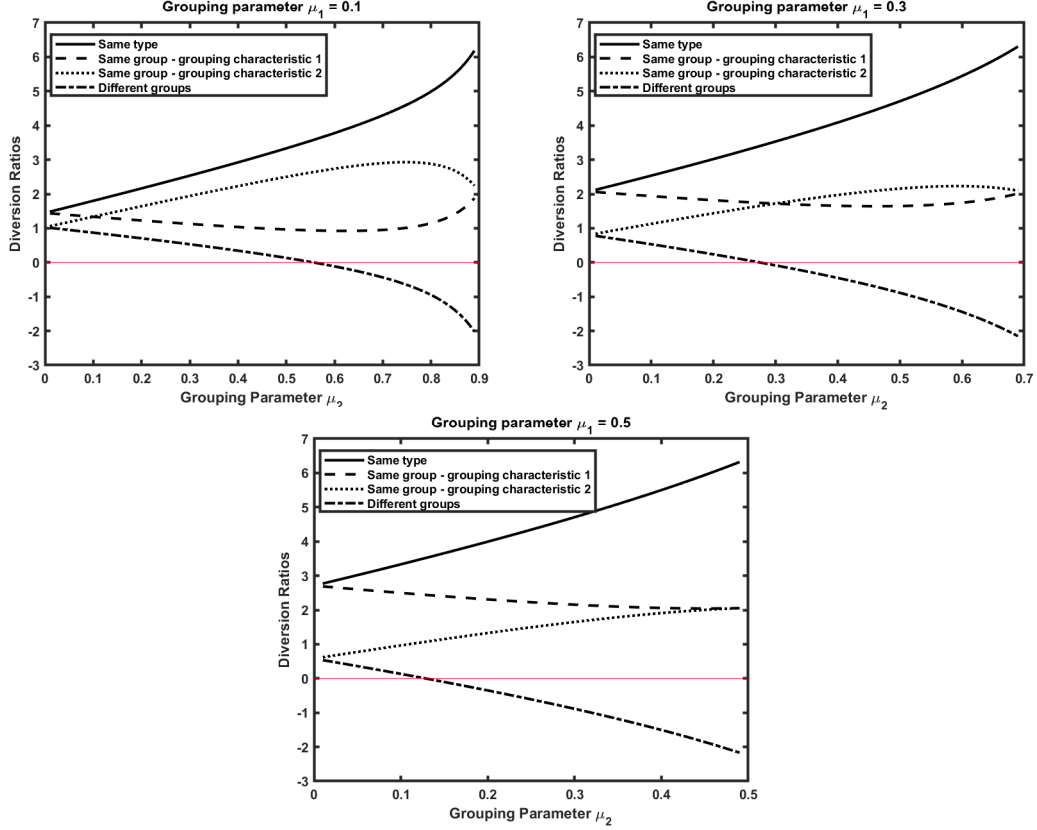
Notes: The figure displays the mean diversion ratio between products in the IPDL model as a function of μ_2 while keeping the value of μ_1 constant. The figure is based on $1,000$ random samples of markets with 45 products. The red horizontal line corresponds to the threshold between complementarity and substitutability in demand.

Complementarity   The empirical literature has mainly used two definitions of complementarity (Berry et al., 2014). Products j and k are complements in demand if the cross-price derivative of market share $\partial \sigma_j(\boldsymbol{\delta}_t)/\partial p_{kt}$ is negative. They are complements in utility if the cross-derivative of utility $\partial u(\mathbf{s}_t)/\partial s_{jt}\partial s_{kt}$ is positive.[7]

By construction, products are always substitutes in utility in the IPDL model. As shown in Figure 1, however, they can be complements in demand, depending on the value of the grouping parameters and the grouping structure.

---

[7]These definitions apply to differentiable demand functions and continuously differentiable utility functions.

Whether two products are complements or substitutes in demand depends on their relationship to other products (Samuelson, 1974). Assume that all products are substitutes in utility. Then, an increase in the price of product 1, for example, has two opposite effects on the market share of product 2. There is a direct substitution effect that increases the market share of product 2 as the market share of product 1 decreases. Note that the market shares of all the other products, including the outside good, also increase following the increase in the price of product 1. There is also an indirect substitution effect via all the products other than products 1 and 2, including the outside good: substitution between product 2 and these products implies that an increase in the market shares of these products causes the market share of product 2 to decrease. If the indirect effect is larger than the direct effect, then an increase in the price of product 1 leads to a decrease in the market shares of both products 1 and 2, making these products complements in demand, even though they are substitutes in utility.

Ogaki (1990) presents a method for computing the direct substitution effect between products 1 and 2 from the estimates of the price derivatives of market shares. He shows that the direct effect can be obtained by removing the effect of the other products (all the products other than products 1 and 2, including the outside good), that is, by considering a change in the market share of product 2 while keeping those of the other products constant. When the direct (resp., indirect) substitution effect is positive, the products are called direct (resp., indirect) substitutes; when it is negative, they are called direct (resp., indirect) complements. See Appendix B.5 for further details.

In the IPDL model, products are necessarily direct substitutes. However, they may be indirect substitutes or complements depending on the value of the grouping parameters and the grouping structure. Therefore, in the IPDL model, complementarity in demand is necessarily due to a negative indirect substitution effect that is larger than a positive direct substitution effect.

To get further intuition on the mechanism generating complementarity in demand in the IPDL model, we consider an example with $J = 3$ products and one outside good. Market shares are equal to $s_{0t} = 1/2$ and $s_{1t} = s_{2t} = s_{3t} = 1/6$ and $\alpha = 1$. Products are grouped according to two grouping characteristics: the

grouping is $\{1\}, \{2, 3\}$ for the first characteristic and $\{1, 2\}, \{3\}$ for the second characteristic. This grouping structure induces products 1 and 2 as well as products 2 and 3 to be substitutes in demand. However, depending on the values of the grouping parameters, products 1 and 3 may be substitutes or complements in demand.

In particular, with $\mu_1 = 1/4$ and $\mu_2 = 1/3$, the direct effect (which equals 0.0976) is larger than the indirect effect (which equals -0.0770), which makes products 1 and 3 substitutes in demand with a cross-price derivative of market share between products 1 and 3 equal to 0.0205. In contrast, with $\mu_1 = 3/5$ and $\mu_2 = 1/3$, the indirect effect (which equals -0.1194) is larger than the direct effect (which equals 0.1087), which makes products 1 and 3 complements in demand, with a cross-price derivative of market share between products 1 and 3 equal to -0.0107. The intuition is that a higher μ_1 (from 1/4 to 3/5) makes products 2 and 3 more substitutable as they belong to the same group for the first grouping characteristic, which translates into a larger indirect effect. See Proposition 3 in Appendix B.3 for details.

3.2 Experiments with Simulated Data

The IPDL model is appealing because it generalizes the nested logit model while retaining its computational simplicity. To highlight the advantages of the IPDL model, we consider three Monte Carlo experiments. The experiments have three main goals: (i) to assess the ability of the IPDL model to approximate the true patterns of substitution and implied markups under different models; (ii) to compare the IPDL model to the state-of-the-art RCL model; and (iii) to check that the IPDL model does not generate complementarity in demand when there is none in the data. We assess approximations and compare models in terms of estimated diversion ratios and implied markups using the Mean Squared Error (MSE). We also report the bias and standard error (S.E.) of these estimates.

For each experiment, we generate 50 datasets consisting of $T = 200$ independent markets with $J = 45$ products, where markets exhibit segmentation according to two grouping characteristics forming four product types. We generate a fully

17

structural model of supply and demand, where the supply side is a static price competition model with five multi-product firms, each with nine products. This allows us to compare models in terms of substitution patterns and markups. See Appendix C for details.

Experiment 1: Data from the IPDL Model   We generate data from the IPDL model and fit the two possible nested logit models and the RCL model with independent normal random coefficients on dummies for groups. This experiment helps assess the bias that results from imposing a hierarchical grouping structure when the true grouping structure is non-hierarchical. It also allows us to check whether the IPDL model allows substitution patterns that the RCL model cannot accommodate. We simulate four IPDL models, varying the values of the grouping parameters such that complementarity in demand occurs in the last two models but not in the first two ones.

We present the results in Table 1. Column (1) shows the true diversion ratios between products of the same type, products that belong to the same group only according to the first grouping characteristic, products that belong to the same group only according to the second grouping characteristic, and products that do not belong to the same group according to either grouping characteristic, as well as the true markups. Columns (2) to (5) compare the estimates of these diversion ratios and markup from the different models we fit in terms of MSE, bias, and standard error. Column (2) provides results for the IPDL model. Column (3) provides results for the nested logit model where the first grouping characteristic defines nests and the second grouping characteristic defines subnests. Column (4) provides results for the nested logit model where the second grouping characteristic defines nests and the first grouping characteristic defines subnests. Finally, Column (5) provides results for the RCL model.

We find first that the correctly specified IPDL model produces estimates that are very close to the truth and with small standard errors. Second, the two nested logit models lead to biased estimates with relatively small standard errors. The bias and standard errors increase when the IPDL model exhibits complementarity in demand since the nested logit models shrink the negative diversion ratios towards zero. This

means that imposing a hierarchical grouping structure can substantially affect the estimated patterns of substitution and implied markups. Third, the RCL model also leads to substantially biased estimates with relatively small standard errors. These biases are even larger when the IPDL model exhibits complementarity in demand. This shows that the IPDL model can produce patterns of substitution and implied markups that the RCL model may fail to approximate, even when the IPDL model does not produce complementarity in demand.

Table 1: Simulation Results when the DGP is the IPDL Model

	(1)	(2)			(3)			(4)			(5)		
	True	IPDL Model			NL Model 1			NL Model 2			RCL Model		
		Bias	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE
Diversion Ratios	DGP: IPDL Model with $\mu_1 = 0.10$ and $\mu_2 = 0.10$												
Same Product Type	1.280	-0.005	0.065	0.004	-0.045	0.074	0.008	-0.270	0.177	0.105	-0.649	0.057	0.424
Same group - grouping characteristic 1	0.867	0.021	0.066	0.005	-0.270	0.105	0.084	-0.344	0.010	0.118	-0.246	0.038	0.062
Same group - grouping characteristic 2	0.878	0.020	0.066	0.005	-0.338	0.011	0.114	0.003	0.090	0.008	-0.267	0.025	0.072
Different groups	0.420	0.031	0.019	0.001	0.133	0.011	0.018	0.108	0.013	0.012	0.181	0.015	0.033
Markups	37.29	0.052	1.029	1.062	-0.405	1.020	1.204	-0.553	1.067	1.444	-0.613	1.033	1.443
Diversion Ratios	DGP: IPDL Model with $\mu_1 = 0.15$ and $\mu_2 = 0.20$												
Same Product Type	1.782	-0.007	0.067	0.005	-0.079	0.084	0.013	-0.180	0.200	0.072	-1.250	0.096	1.572
Same group - grouping characteristic 1	0.908	0.034	0.072	0.006	-0.572	0.121	0.342	-0.536	0.009	0.288	-0.380	0.085	0.152
Same group - grouping characteristic 2	1.112	0.040	0.072	0.007	-0.710	0.011	0.505	-0.221	0.107	0.060	-0.618	0.034	0.383
Different groups	0.138	0.037	0.025	0.002	0.283	0.010	0.080	0.245	0.011	0.060	0.351	0.020	0.124
Markups	33.04	0.081	0.918	0.850	-0.751	0.896	1.366	-0.574	0.954	1.240	-1.125	0.975	2.216
Diversion Ratios	DGP: IPDL Model with $\mu_1 = 0.20$ and $\mu_2 = 0.30$												
Same Product Type	2.399	-0.016	0.069	0.005	-0.095	0.098	0.019	-0.014	0.226	0.051	-1.944	0.116	3.792
Same group - grouping characteristic 1	1.018	0.042	0.077	0.008	-0.887	0.143	0.807	-0.762	0.008	0.580	-0.569	0.093	0.332
Same group - grouping characteristic 2	1.394	0.058	0.078	0.010	-1.103	0.011	1.216	-0.507	0.126	0.273	-0.987	0.041	0.977
Different groups	-0.162	0.014	0.032	0.001	0.475	0.010	0.226	0.431	0.008	0.186	0.563	0.020	0.317
Markups	28.12	0.075	0.788	0.626	-0.984	0.762	1.549	-0.523	0.826	0.956	-1.572	0.871	3.231
Diversion Ratios	DGP: IPDL Model with $\mu_1 = 0.25$ and $\mu_2 = 0.40$												
Same Product Type	3.104	0.002	0.067	0.005	-0.039	0.118	0.016	0.302	0.249	0.153	-2.646	0.222	7.050
Same group - grouping characteristic 1	1.184	0.063	0.080	0.010	-1.178	0.172	1.416	-1.015	0.009	1.030	-0.772	0.157	0.621
Same group - grouping characteristic 2	1.698	0.077	0.082	0.013	-1.493	0.013	2.228	-0.878	0.144	0.792	-1.334	0.108	1.791
Different groups	-0.463	-0.090	0.041	0.010	0.691	0.013	0.478	0.647	0.001	0.418	0.793	0.032	0.630
Markups	22.26	0.036	0.630	0.399	-1.055	0.616	1.493	-0.379	0.679	0.605	-1.839	0.717	3.895

Notes: Summary statistics across 50 Monte Carlo replications. In the first two DGPs, there is no complementarity in demand. In the last two DGPs, 21% of the pairs of products exhibit complementarity in demand.

Experiment 2: Data from Bresnahan et al. (1997)'s PDL Model We generate data from the PDL model and fit the IPDL model and the RCL model with independent normal random coefficients on dummies for groups. This second experiment helps assess the performance of the IPDL model in predicting the patterns of substitutions and markups generated by another model of segmentation. It also allows us to compare the IPDL model to the RCL model when both are misspecified. In addition, the experiment tests whether the IPDL model may wrongly generate complementarity in demand in a case where products are substitutes in demand. We

simulate four PDL models, varying the values of the grouping parameters μ_1 and μ_2 that control substitution between products.

Table 2 presents the results. Column (1) provides the true diversion ratios between products and the true markups. Columns (2) to (4) compare the estimates of these diversion ratios and the markup from the PDL model, the IPDL model, and the RCL model, respectively.

Table 2: Simulation Results when the DGP is the PDL Model

	(1)	(2)			(3)			(4)		
	True	PDL Model			IPDL Model			RCL Model		
	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE	Bias	S.E.	MSE
Diversion Ratios	DGP: PDL Model with $\mu_1 = 0.50$ and $\mu_2 = 0.30$									
Same Product Type	2.932	-0.006	0.071	0.005	-0.257	0.077	0.072	-1.405	0.415	2.146
Same group - grouping characteristic 1	1.813	0.004	0.043	0.002	0.059	0.097	0.013	-0.422	0.249	0.240
Same group - grouping characteristic 2	1.627	-0.005	0.025	0.001	-0.116	0.010	0.023	-0.362	0.206	0.174
Different groups	0.594	0.002	0.028	0.001	-0.136	0.056	0.022	0.570	0.235	0.381
Markups	26.05	-0.030	0.732	0.536	0.002	0.689	0.475	-0.676	0.854	1.186
Diversion Ratios	DGP: PDL Model with $\mu_1 = 0.50$ and $\mu_2 = 0.50$									
Same Product Type	2.716	0.017	0.075	0.006	-0.144	0.063	0.025	-1.199	0.354	1.562
Same group - grouping characteristic 1	1.711	-0.006	0.101	0.010	-0.025	0.083	0.008	-0.281	0.171	0.108
Same group - grouping characteristic 2	1.705	0.014	0.099	0.010	-0.007	0.083	0.007	-0.367	0.142	0.155
Different groups	0.730	-0.011	0.034	0.001	-0.127	0.047	0.018	0.544	0.208	0.340
Markups	29.25	-0.108	0.870	0.769	-0.169	0.819	0.700	-0.707	0.879	1.272
Diversion Ratios	DGP: PDL Model with $\mu_1 = 0.50$ and $\mu_2 = 0.70$									
Same Product Type	2.469	0.012	0.079	0.006	-0.124	0.058	0.019	-0.896	0.516	1.068
Same group - grouping characteristic 1	1.931	0.015	0.112	0.013	-0.163	0.077	0.033	-0.473	0.173	0.254
Same group - grouping characteristic 2	1.463	-0.009	0.093	0.009	0.123	0.075	0.021	-0.039	0.130	0.018
Different groups	0.912	-0.008	0.041	0.002	-0.047	0.043	0.004	0.461	0.210	0.257
Markups	32.93	-0.007	0.939	0.882	-0.074	0.907	0.829	-0.762	0.915	1.419
Diversion Ratios	DGP: PDL Model with $\mu_1 = 0.50$ and $\mu_2 = 0.90$									
Same Product Type	2.355	0.019	0.068	0.005	-0.144	0.059	0.024	-0.776	0.250	0.664
Same group - grouping characteristic 1	2.376	-0.003	0.075	0.006	-0.282	0.076	0.085	-0.945	0.179	0.926
Same group - grouping characteristic 2	1.030	0.008	0.056	0.003	0.196	0.073	0.044	0.490	0.188	0.275
Different groups	0.982	-0.010	0.039	0.002	0.040	0.045	0.004	0.389	0.184	0.186
Markups	34.60	0.011	0.952	0.907	-0.125	0.911	0.845	-1.401	0.878	2.734

Notes: Summary statistics across 50 Monte Carlo replications.

The estimates from the correctly specified PDL model are very close to the true values, with small standard errors. The estimates from the IPDL model are also close to the true values but not as close as the correctly specified PDL model, with small standard errors. We also find that the IPDL model does not wrongly produce complementarity in demand. Finally, as in the previous experiment, the RCL model leads to substantially biased estimates, and with larger standard errors. This experiment provides thus a case where the IPDL model outperforms the RCL model.

Experiment 3: Data from the RCL Model   In this experiment, we generate data from the RCL model with independent normal random coefficients on dummies for groups. This third experiment helps assess the ability of the IPDL model to approximate the patterns of substitution and implied markups generated by the popular RCL model. It also allows us to check whether the IPDL model wrongly generates complementarity in demand in another case where products are substitutes in demand. We simulate four RCL models, varying the values of the standard deviations of the normal random coefficients RC_1 and RC_2.

Table 3: Simulation Results when the DGP is the RCL Model

	True	RCL Model			IPDL Model		
		Bias	S.E.	MSE	Bias	S.E.	MSE
Diversion Ratios	DGP: RCL Model with $RC_1 = 0.50$ and $RC_2 = 1.00$						
Same Product Type	2.267	0.001	0.021	0.001	-0.002	0.019	0.001
Same group - grouping characteristic 1	2.127	0.005	0.035	0.001	0.022	0.038	0.002
Same group - grouping characteristic 2	2.142	-0.004	0.032	0.001	0.035	0.028	0.002
Different groups	1.929	0.001	0.028	0.001	0.013	0.028	0.001
Markups	30.16	0.109	0.417	0.186	0.025	0.420	0.177
Diversion Ratios	DGP: RCL Model with $RC_1 = 1.00$ and $RC_2 = 2.00$						
Same Product Type	2.512	-0.001	0.019	0.000	-0.067	0.018	0.005
Same group - grouping characteristic 1	1.947	0.003	0.035	0.001	0.247	0.041	0.063
Same group - grouping characteristic 2	2.058	-0.001	0.034	0.001	0.095	0.026	0.010
Different groups	1.414	0.002	0.017	0.000	0.015	0.031	0.001
Markups	30.29	0.112	0.422	0.190	-0.105	0.437	0.202
Diversion Ratios	DGP: RCL Model with $RC_1 = 1.50$ and $RC_2 = 3.00$						
Same Product Type	2.795	-0.002	0.024	0.001	-0.162	0.017	0.027
Same group - grouping characteristic 1	1.765	0.001	0.026	0.001	0.598	0.042	0.359
Same group - grouping characteristic 2	1.912	0.001	0.037	0.001	0.194	0.025	0.038
Different groups	1.007	0.002	0.017	0.000	-0.130	0.032	0.018
Markups	30.43	0.110	0.428	0.195	-0.254	0.475	0.289
Diversion Ratios	DGP: RCL Model with $RC_1 = 2.00$ and $RC_2 = 4.00$						
Same Product Type	3.053	-0.001	0.016	0.000	-0.264	0.016	0.070
Same group - grouping characteristic 1	1.604	0.002	0.018	0.000	0.975	0.042	0.953
Same group - grouping characteristic 2	1.744	-0.001	0.024	0.001	0.314	0.024	0.010
Different groups	0.735	0.001	0.009	0.000	-0.354	0.033	0.127
Markups	30.54	0.116	0.428	0.196	-0.412	0.529	0.449

*Notes:*Summary statistics across 50 Monte Carlo replications.

Table 3 presents the results. Column (1) provides the true diversion ratios between products and the true markups. Columns (2) and (3) compare the estimates of these diversion ratios and markups from the RCL model and the IPDL model, respectively.

We find first that the correctly specified RCL model produces estimates very close to the true values and with small standard errors. Second, the IPDL model generates estimates reasonably close to the true values with small standard errors. The biases increase with the standard deviations of the normal random coefficients, meaning it becomes harder to approximate the RCL model as it deviates more from

the logit model. However, the biases remain small, except for the diversion ratios between products sharing one group on the first grouping characteristic. These simulations thus show that the IPDL model can approximate the rich substitution patterns of the RCL model, at least when the standard deviations of the normal random coefficients are not too large. We find, once again, that the IPDL model does not wrongly produce complementarity in demand.

3.3 The IPDL Model versus Other Models

GEV Models   The GEV family of models encompasses all additive random utility models in which the random utility terms follow a multivariate extreme value distribution (McFadden, 1978; Anderson et al., 1992; Fosgerau et al., 2013).

Except for the logit model, the nested logit model is the simplest and most popular GEV model. The nested logit model market share function is closed form. It partitions the choice set into nests, nests into subnests, etc. This hierarchical grouping structure imposes restrictions on substitution patterns.[8] Despite these restrictions, the nested logit model is commonly used in applied work with market-level data due to its computational simplicity. Indeed, it is estimated by linear instrumental variables regression, using that its inverse market share function has a linear-in-parameters, closed-form expression (Berry, 1994).

Bresnahan et al. (1997) propose the PDL model, a GEV model that accommodates richer substitution patterns than the nested logit model by allowing a non-hierarchical grouping structure. Like the nested logit model, the PDL market share function is in closed form. However, the PDL model inverse market share function is not in closed form. The PDL model cannot, therefore, be estimated simply by linear instrumental variable regression. Instead, estimation of the PDL requires minimizing a non-linear, non-convex generalized method of moments objective, which must be computed by inverting the PDL market share function numerically.

---

[8]In the three-level nested logit model, for example, there are only three different cross-price elasticities per product: one for products within the same subnest, another lower cross-price elasticity for products from different subnests within the same nest, and an even lower cross-price elasticity for products from different nests. The cross-price elasticities are necessarily positive, i.e., all products are substitutes in demand.

The IPDL model generalizes the nested logit model by allowing the same non-hierarchical grouping structure as the PDL model. In contrast to the PDL model, the IPDL model can be estimated by linear instrumental variable regression but its market share function is not generally in closed form. In contrast to the nested logit and the PDL models, the IPDL model allows complementarity in demand. In summary, the IPDL model allows richer substitution patterns than the nested logit model while retaining its simplicity of estimation. It allows substitution patterns not accommodated by the PDL model while being simpler to estimate with market-level data.

A range of GEV models has been proposed using various grouping structures other than that of the PDL model. Prominent examples include the ordered logit (Small, 1987), the paired combinatorial logit (Koppelman and Wen, 2000), the flexible coefficient multinomial logit (Davis and Schiraldi, 2014), and the ordered nested logit (Grigolon, 2021). It is straightforward to extend the IPDL model to use the grouping structures of these GEV models. All such models would still have a linear-in-parameter inverse market share function and would, therefore, still be estimated by linear instrumental variable regression. This is, for example, part of the strategy proposed by Monardo (2021), who builds an inverse market share model, which, like Koppelman and Wen (2000) and Davis and Schiraldi (2014)'s models, employs a grouping structure with a group for each pair of products. See Fosgerau et al. (2021) and the supplement for further details. Finally, using our setting, Hortaçsu et al. (2020) propose a method to estimate the grouping structure from the data.

The generalized nested logit (Wen and Koppelman, 2001) and the cross-nested logit (Vovsha, 1997; Ben-Akiva and Bierlaire, 1999) go further and allow partial group membership, with an additional set of parameters controlling for the degree of group memberships. It is possible to extend the IPDL model via a similar construction. In such extensions, the inverse market share function would not be linear in the parameters controlling for the degree of group membership. Estimation would, therefore, require more complex non-linear instrumental variable regression.
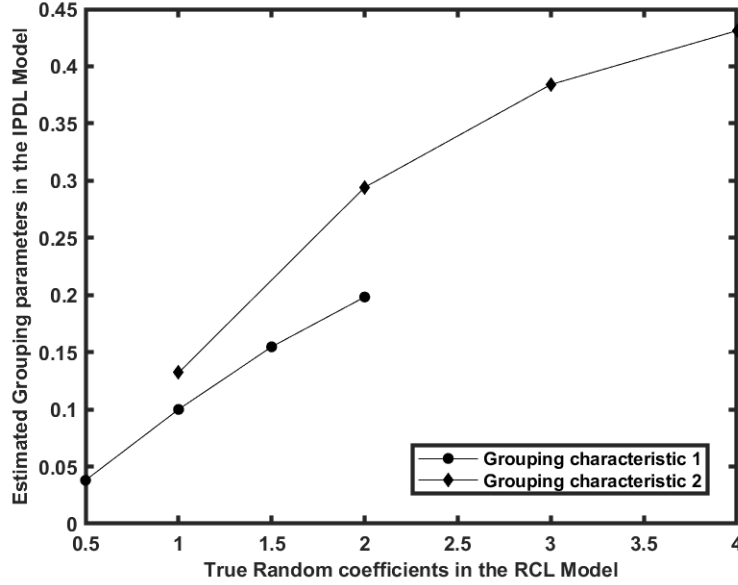
The Random Coefficient Logit Model   The RCL model has been the state-of-the-art model in the literature since the seminal paper by Berry et al. (1995). The RCL model extends the logit model by incorporating unobserved preference heterogeneity through the specification of random coefficients on product characteristics, including prices. Most papers that estimate the RCL model with market-level data assume independent normal random coefficients and use the estimation algorithm proposed by Berry et al. (1995), known as the BLP method. The RCL model accommodates rich substitution patterns determined by how close products are in the space of product characteristics that receive a random coefficient. However, the BLP method is computationally demanding as it involves a non-linear, non-convex optimization problem and the simulation and numerical inversion of the market share function.

Furthermore, when the RCL model has random coefficients on dummies for groups, it produces substitution determined by the random coefficients distribution and by the group memberships. Similarly, the IPDL model delivers substitution determined by the grouping structure and the corresponding grouping parameters.

As shown by Cardell (1997) and further studied by Galichon (2021), the (two-level) nested logit model is an RCL model for which the dummy variables that form the grouping structure receive a random coefficient with a specific distribution. This observation motivates the open question of whether an IPDL model is equivalent to some RCL model. We can immediately rule out IPDL models exhibiting complementarity in demand since products can only be substitutes in demand in the RCL model.

Furthermore, the RCL model, with negative coefficients on prices, satisfies the condition that the higher order partial derivatives of the market share function σ_i ($i \in \mathcal{J}$) with respect to any set of distinct prices other than p_i are non-negative (see Theorem 3.1 Anderson et al., 1992, page 67). This condition rules out complementarity in demand. By contrast, the IPDL model does not necessarily satisfy this condition, even when there is no complementarity in demand. This means that the IPDL model allows behavior that cannot be accommodated by any RCL model, even when all products are substitutes in demand. See Appendix B.4 for details.

Figure 2: Relationship between Grouping Parameters in the IPDL model and Random Coefficients in the RCL model

To gain further insights on the comparison between the IPDL model and the RCL model with random coefficients on dummies for groups, we consider again Experiment 3, where we simulate four RCL models with independent normal random coefficients on two dummies for groups, varying the values of the random coefficients, i.e., the standard deviations of the normal random coefficients. Figure 2 shows the mean of the estimated grouping parameters in the misspecified IPDL models against the true value of the random coefficients in the RCL models for the four designs.[9] As expected, we find an increasing relationship between the estimated grouping parameters and the true random coefficients: higher random coefficients in the RCL model means greater deviations from the logit model, and the same does higher grouping parameters in the IPDL model. It suggests, as mentioned above, that the grouping parameters in the IPDL model are consumer heterogeneity parameters, like the random coefficients in the RCL model and the grouping parameters in the nested logit model. This interpretation of the grouping parameters is consistent with the interpretation of the IPDL model as a model of

---

[9]Grigolon and Verboven (2014) propose a similar figure and discussion for the nested logit model.

utility-maximizing, heterogeneous consumers. Furthermore, the grouping parameters are increasing in the random coefficients at a decreasing rate such that their values remain low enough to be consistent with the assumptions of the IPDL model (i.e., with two grouping characteristics, $\mu_1 \geq 0$, $\mu_2 \geq 0$, and $\mu_1 + \mu_2 < 1$).

Models with Complementarity   The literature has allowed for complementarity in two main ways. The first strand of empirical literature incorporates complementarity in demand through micro-founded demand systems. Prominent examples include the AIDS model of Deaton and Muellbauer (1980), the EASI model of Lewbel and Pendakur (2009), and the linear demand model (see, e.g., Pinkse and Slade, 2004; Thomassen et al., 2017; Lewbel and Nesheim, 2019). The IPDL model belongs to this strand. The IPDL model, however, differs from these papers regarding how unobservables enter the model. In all these models, the unobservables enter in a very restrictive way: the unobservables of a given product affect only its own demand, whereas the IPDL model, via the terms ξ_{jt}, allows the unobservables of a given product to affect both its own demand (market share) and those of all its competing products.

Another strand of literature incorporates complementarity by building demands for baskets of products. Prominent examples include Gentzkow (2007), Iaria and Wang (2020), and Ershov et al. (2021), who directly extend the RCL model to allow each consumer to choose among baskets of products rather than products alone. In their models, complementarity arises from positive demand synergies: products are complements in utility if the utility that a consumer derives from consuming the basket is higher than the sum of the utilities she derives from consuming the products separately. Products can also be complements in demand. Another example is Iaria and Wang (2021), who, similarly to us, build a demand model based on a grouping structure, which can be inverted to obtain an inverse market share function that can easily be estimated even with large choice sets. Our paper thus differs from these papers regarding the channel through which products are complements in demand. Further, we only need to observe purchases at the product level, whereas these approaches rely on purchases at the basket level.

Summary   We illustrate the relationship between the logit, nested logit, RCL and IPDL models in Figure 3. As is well known, the logit model is a special case of the nested logit model. Besides, all nested logit models are both IPDL models and RCL models. Otherwise, the sets of RCL models and IPDL models are generally different. We cannot exclude that some IPDL models are equivalent to some RCL models. Finding general conditions under which this is the case, as for the nested logit model, is a hard mathematical problem that we leave for future research.

Figure 3: Relationships between logit, nested logit, RCL, and IPDL models
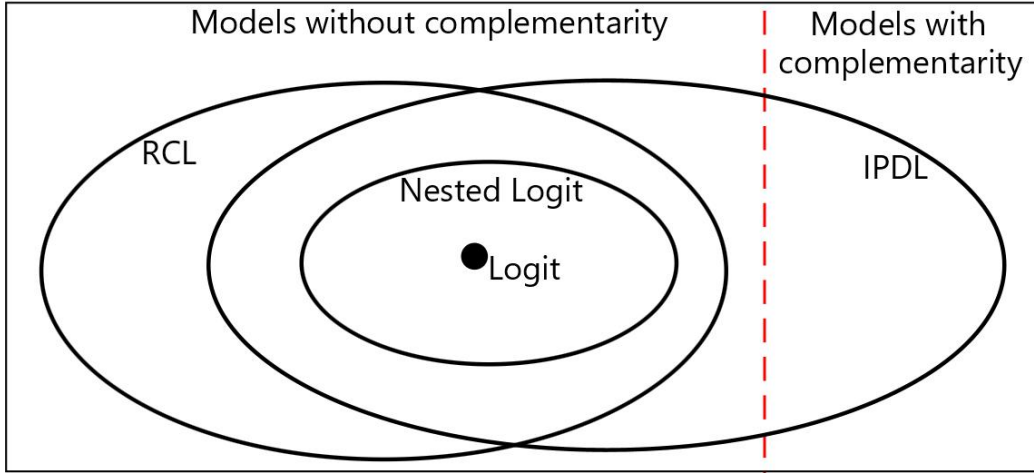
4   Empirical Application

In this section, we use the IPDL model to estimate the demand in the ready-to-eat cereals market, which has been studied extensively (Nevo, 2000, 2001; Conlon et al., 2021; Michel et al., 2022). We have three main goals: (i) to show how the IPDL model works with a well-known dataset; (ii) to investigate the computational performances of the IPDL model when there are many markets or many products; (iii) to compare the IPDL model to the RCL model estimated using the methodology developed by BLP in terms of computational performance and goodness-of-fit. Details for this section are provided in Appendix D.

4.1 Data

Data Sources   We use data from the Dominick's Dataset, which is made publicly 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 Universal Product Code (UPC) level for all Dominick's Finer Foods 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 daily store traffic.

For our analysis, we consider the ready-to-eat cereals category during the period 1991 –1996. We use data from 25 Dominick's stores, and we aggregate UPCs into what we call brand-name cereals (e.g., Kellogg's Special K). We select 45 brand-name cereals from 6 national manufacturers (General Mills, Kellogg's, Nabisco, Post, Quaker and Ralston), representing around 75% of each manufacturer's total sales on the period.[10] We define three market segments, namely Adults, Kids and All-family, according to the classification provided by the website cerealfacts.org.

Prices are retail prices calculated as the volume-weighted average price per ounce of the UPCs that form the product, deflated by the monthly Consumer Price Index for All Urban Consumers in the Chicago-Naperville-Elgin area from the U.S. Bureau of Labor Statistics. We compute the potential market size by multiplying the total number of persons in a market by the monthly per capita consumption of cereals. We compute the total volume of a product sold in a market, which we divide by the potential market size to obtain the product's market share. The market share of the outside good is then the difference between one and the sum of the 45 products' market shares.

We supplement Dominick's Dataset with information on the nutrient content (fiber, sugar, and calories) of the cereals from the USDA Nutrient Database for Standard Reference (release SR11, year 1996) and on the type of grains (rice, wheat, corn, and oats) using manufacturers' websites and different websites collecting nutritional information. We also use monthly input prices from the websites index-mundi.com (corn, rice, sugar, and wheat) and macrotrends.net (oats) to construct

---

[10]Only package sizes between 10 and 32 ounces are included. The 45 brand-name cereals account for around 58% of the national market (see, e.g., Corts, 1996).

cost-based instruments.

Descriptive Statistics Table 4 presents descriptive statistics on market shares and retail prices of cereals by brand and market segment. Kellogg's and General Mills are the largest two brands and are active in all market segments. Market segments have about equal market shares, and cereals for kids have higher prices on average. Taken together, Kellogg's and General Mills account for around 73 percent of the market, excluding the outside good. Furthermore, cereals for kids tend to be more expensive than their competitors, cereals for adults tend to be cheaper, General Mills and Ralston set higher prices, and Quaker lower prices.

Table 4: Shares and prices by brand and market segment

	All-family		Adults		Kids		Total	
	shares	prices	shares	prices	shares	prices	shares	prices
General Mills	3.39	20.12	2.04	20.18	3.23	21.04	8.66	20.48
Kellogg's	1.29	17.04	6.10	16.84	6.34	18.39	13.73	17.57
Nabisco	–	–	0.69	18.07	–	–	0.69	18.07
Post	0.84	16.44	1.63	15.98	0.97	21.91	3.44	17.77
Quaker	2.09	15.79	1.19	14.45	–	–	3.28	15.30
Ralston	0.75	20.91	–	–	0.19	24.55	0.94	21.65
Total	8.36	18.26	11.65	17.13	10.73	19.61	30.74	18.69
Outside good							69.26	

Notes: Shares and prices refer to average (across markets) market shares in percent and retail prices (in cents) per ounce, respectively

Table 5 shows the average nutrient content of the cereals by brand and market segment. Cereals offered by Nabisco contain, on average, less sugar, fat, and sodium and more fat and protein than those of its competitors. In contrast, cereals offered by General Mills and Post are, on average, more sugary, cereals offered by Quaker have more fat and are rather sugary, and those offered by Nabisco and Ralston have less fat. Cereals offered by Kellogg's are rather sugary. Furthermore, cereals for kids contain more sugar than cereals of the other segments; they also contain less fiber and protein. By contrast, cereals for adults tend to have less sugar and sodium but more fiber.

Table 5: Nutrients content by brand and market segment

	Sugar g/ounce	Fiber g/ounce	Fat g/ounce	Protein g/ounce	Sodium mg/ounce	#
Brands						
General Mills	8.50	1.68	0.25	0.59	194.82	10
Kellogg's	7.85	1.32	0.23	0.59	149.59	15
Nabisco	0.24	3.25	0.057	0.83	1.98	2
Post	8.68	1.91	0.28	0.57	170.48	9
Quaker	7.85	1.50	0.76	0.69	143.56	5
Ralston	5.02	1.10	0.11	0.54	239.91	4
Segments						
All-family	7.04	1.64	0.23	0.62	195.68	12
Adults	5.17	2.19	0.32	0.77	135.56	19
Kids	11.29	0.77	0.29	0.36	177.43	14
All	7.57	1.60	0.29	0.60	164.62	45

Notes: Nutrient content refers to (unweighted) averages across cereals, by brand and market segment. Column # gives the number of products by brand and market segment.

Overall, we can view the brands and market segments as proxying, at least partially, the nutrient content of the cereals as well as their prices. As a result, consider an IPDL model that groups cereals according to the brands and market segments. This IPDL model has substitution patterns depending on this grouping structure and, thus, indirectly on the nutrient content and prices.

4.2 Specification and Identification

Specification We specify an IPDL model with two grouping characteristics: i) the market segment the cereals belong to (F for All-family, A for Adults, and K for Kids), and ii) the brand the cereals belong to (G for General Mills, K for Kellogg's, N for Nabisco, P for Post, Q for Quaker, and R for Ralston). We estimate this IPDL model using the linear instrumental variables regression (9) with $D = 2$ grouping characteristics, where $\mathbf{x}_{jt}$ includes a constant and a yearly trend, and where $d_1 = \{F, A, K\}$, and $d_2 = \{G, K, N, P, Q, R\}$.

We compare the IPDL model to the RCL model. We specify an RCL model with independent normal random coefficients on a constant and on the dummies for the groups *F*, *K*, *N-R*, and *P-Q*.[11] We estimate this RCL model using the methodology

---

[11]For the sake of parsimony, we divide brands into three groups according to their popularity

developed by Berry et al. (1995) and implementing the best practices as advocated by Conlon and Gortmaker (2020).

We estimate two specifications of the IPDL and RCL models, varying the definition of products and markets. In the first specification (large T), we define a product as a brand-name cereal and a market as a store-month pair. As a result, the sample covers $J = 45$ in $T = 1,675$ markets. In the second specification (large J), we define a product as a brand-name cereal/store pair and a market as a month. As a result, the sample covers $J = 1,125$ in $T = 67$ markets. In both specifications, we include fixed effects for products and months. We further include fixed effects for stores in the first specification.

Identification To identify the substitution patterns, we rely on two sets of instruments. The first set uses a second-order polynomial of cost shifters and continuous exogenous characteristics. The cost shifters are the input prices (sugar, corn, oats, rice, and wheat) multiplied by the corresponding characteristics so that they vary by brand-name cereals and across time. The product characteristics are sugar, fiber, fat, protein, and sodium content.

The second set consists of BLP-type instruments. Specifically, we use the quadratic version of Gandhi and Houde (2021)'s differentiation IVs. For each exogenous continuous product characteristic x^k and for each pair (i, j) of products, we compute the differences $d_{i,j,t}^k = x_{jt}^k - x_{it}^k$. The differentiation IVs we consider are all quadratic interactions of these differences, summed over products belonging to the same brand, to a competing brand, to the same market segment, and to a competing segment. Furthermore, we compute exogenous price indices, $\hat{p}_{jt}$, as the predicted values from a linear regression of the price variable on the two sets of instruments and a constant. We then compute the differences $d_{i,j,t}^{\hat{p}} = \hat{p}_{jt} - \hat{p}_{it}$, which we interact with itself and the other differences to construct additional differentiation IVs.

For the IPDL models, we use the exogenous price index as an instrument and construct instruments for the two log-share terms as the predicted values from regressions of these terms on the two sets of instruments and a constant.[12] For the

---

measured in terms of market shares: General Mills and Kellogg's, Nabisco and Ralston, and Post and Quaker.

    [12]A potential problem is weak identification, which occurs when instruments are only weakly

RCL models, we first estimate using the exogenous price index, the two sets of instruments, and a constant as instruments. Then, based on these estimates, we compute the optimal instruments and estimate using these instruments.

4.3 Results

Table 6 presents the estimation results. Columns (1) and (2) provide the results for the large T and the large J specifications of the IPDL model. Columns (3) and (4) provide the results for the large T and the large J specifications of the RCL model.

Demand Parameters The estimated parameter on the negative of price (α) has the expected sign and is significantly different from zero in both specifications of the IPDL and the RCL models. For the IPDL model, the estimates of α have the same magnitude in the two specifications. By contrast, for the RCL model, it is significantly higher in the large J specification than in the large T specification.

For both specifications of the IPDL model, the grouping parameters are precisely estimated and satisfy the assumptions of the IPDL model.[13] Furthermore, the grouping parameter for brand name is higher than that for market segment (i.e., $\mu_1 > \mu_2$), which indicates that brand reputation confers more protection from substitution than does the market segment, i.e., cereals of the same brand are more protected from cereals from other brands than cereals of the same market segment are from cereals from different market segments.

For the RCL model, only two and three of the five random coefficients are precisely estimated in the large T and large J specifications, respectively. This is consistent with the literature, which has found that it may be hard in practice to identify random coefficients on dummies when product fixed effects are included in the model (e.g., Conlon and Mortimer, 2013). Moreover, random coefficients do not have the same magnitude in the two specifications.

---

correlated with the endogenous variables. In both specifications, the Sanderson and Windmeijer (2016)'s F-statistics to test whether each endogenous variable is weakly identified are far above 10, the rule-of-thumb usually used for linear instrumental variables regressions, thereby suggesting that instruments are not weak.

[13]i.e., $\mu_1 \geq 0$, $\mu_2 \geq 0$, and $\mu_1 + \mu_2 < 1$. We impose no constraints on the parameters during the estimation.

Table 6: Estimation Results

	(1) IPDL Model Large T	(2) IPDL Model Large J	(3) RCL Model Large T	(4) RCL Model Large J
Prices ($-\alpha$)	-12.848	-12.208	-16.380	-25.513
	(0.3123)	(0.2704)	(1.840)	(0.7372)
Grouping Parameters				
Brand name (μ_1)	0.4699	0.4886	–	–
	(0.0200)	(0.0179)	–	–
Market segment (μ_2)	0.2856	0.2825	–	–
	(0.0222)	(0.0199)	–	–
Random Coefficients				
Constant	–	–	1.023	0.8455
	–	–	(1.024)	(1.416)
Nabisco-Ralston	–	–	0.7566	1.154
	–	–	(1.017)	(0.3121)
Post-Quaker	–	–	0.3978	0.8485
	–	–	(0.2143)	(0.4014)
Kids	–	–	0.8823	4.792
	–	–	(0.5935)	(0.7102)
All-family	–	–	0.1184	2.415
	–	–	(2.037)	(1.580)
Product Fixed Effects	45	1125	45	1125
Month Fixed Effects	11	11	11	11
Store Fixed Effects	24	–	24	–
Time	5sec	2min	~1h50	~2h30
Cross-validated MSE	0.169	0.212	0.347	0.341
Complementarity				
Complements (%)	5.25	5.74	0	0
Mean Diversion Ratios				
Same product type	10.04	0.383	1.28	0.126
Same group - brand name	5.27	0.190	1.01	0.018
Same group - market segment	3.04	0.110	1.17	0.107
Different groups	0.371	0.013	0.883	0.030
Mean Markup	36.51%	38.47%	38.76%	26.75%

Notes: The number of observations is 75,375. Robust standard errors are shown in parentheses. A constant and a yearly trend are included.

Computational time   We compare the computational time of the IPDL and the RCL models, measured as the execution time from the computation of the instruments to the estimation of the model. We find that the IPDL model is much faster to estimate than the RCL model. The IPDL model takes 5 seconds to estimate the large T specification and 2 minutes to estimate the large J specification, against

1h50 and 2h30, respectively, for the RCL model.

Goodness-of-fit   We compare the goodness-of-fit of the IPDL and the RCL models. Like Compiani (2022), we measure goodness-of-fit based on the cross-validated mean squared error (MSE) to evaluate how models perform out-of-sample. In both specifications, we find a lower cross-validated MSE for the IPDL model than for the RCL model, meaning that the IPDL model fits the data better than the RCL model.[14]

Substitution Patterns and Markups   We provide the diversion ratios between cereals averaged across markets and products according to whether they are of the same types, have the same brand name, belong to the same market segment, or do not have the same brand name and do not belong to the same market segment. We find significantly different diversion ratios between the IPDL and the RCL models. For the large T specification, in the IPDL model, diversion ratios are determined by segmentation, whereas in the RCL model, the diversion ratios do not seem to be determined by segmentation. For the large J specification, both the IPDL and RCL models generate diversion ratios depending on segmentation. However, they lead to different qualitative results. For example, the IPDL model predicts that cereals with the same brand name are closer substitutes than cereals belonging to the same market segment, and conversely for the RCL model. Lastly, in both specifications, the IPDL model generates significantly higher substitution between cereals than the RCL model.

We also provide the percentage of pairs of complements in demand in the IPDL model, determined by the percentage of significantly negative diversion ratios. This takes into account that some pairs of cereals exhibiting negative diversion ratios may actually be independent and not complements in demand. We find that the IPDL model generates a small amount of complementarity in demand, as only around 5% of the pairs of products exhibit complementarity in demand. To verify that complementarity in demand is not a model artifact, we compute the diversion

---

[14]We have also tested more usual specifications of the RCL model, where continuous product characteristics (sugar, fiber, and fat) and price receive a random coefficient and are interacted with demographics (income and child). None of these specifications provided a better fit to the data.

ratios using different values of the grouping parameters from the estimated values. We find that parameter values exist such that the model does not exhibit complementarity in demand.

Finally, we provide markups, averaged across markets and products, computed by assuming static oligopolistic price competition between firms. We find that the IPDL model yields markups that have the same magnitude across the two specifications and are in line with the literature (Nevo, 2001; Michel et al., 2022). By contrast, the RCL model yields similar markups as the IPDL model in the large T specification but significantly lower than the IPDL model in the large J specification (because of the large estimates of the parameter α).

5 Concluding Comments

We have introduced the IPDL model, a micro-founded inverse market share model for differentiated products. The IPDL model generalizes the nested logit model to allow richer substitution patterns, including complementarity in demand. Like the nested logit model, the IPDL model can be estimated by linear instrumental variable regression using market-level data, and it is consistent with a model of heterogeneous, utility-maximizing consumers.

Our Monte Carlo experiments show that the IPDL model can reasonably approximate the substitution patterns generated by the workhorse models of the literature. Our empirical application, using a well-known dataset on the ready-to-eat cereals market, shows that the IPDL model fits the data better than a similar RCL model while being much faster to estimate. With the IPDL model, we find evidence of complementarity in demand due to the indirect substitution effect, a result that would not be possible with the RCL model. These results suggest that the IPDL model can be useful for describing markets that exhibit segmentation.

This paper opens several avenues for future work. First, it would be interesting to see applications of the IPDL model to different markets and economic issues. The IPDL model may lead to qualitatively different conclusions from the workhorse models of the literature, particularly when there is complementarity in demand. Second, it seems worthwhile to extend the IPDL model to allow for un-

observed heterogeneity in preferences through random coefficients, in analogy with what has been done with the logit and nested logit models. Third, a natural next step would be to develop an estimation method for the IPDL model using individual-level data rather than market-level data. Finally, it would be interesting to incorporate forward-looking consumer behavior in the IPDL model and to develop a corresponding estimation procedure.

Appendix

A   Mathematical Appendix

Notation   We use italics for scalar variables and real-valued functions, boldface for vectors, matrices and vector-valued functions, and calligraphic for sets. $\mathbb{R}_+$ is the set of non-negative real numbers, $\mathbb{R}_{++}$ is the set of positive real numbers, and $\mathbb{R}_{++}^{J+1} = (0, \infty)^{J+1}$. As default, vectors are column vectors: $\mathbf{s} = (s_0, \ldots, s_J)^\mathsf{T} \in \mathbb{R}^{J+1}$.

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

Let $\mathbf{G} = (G_0, \ldots, G_J) : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a vector function composed of functions $G_j : \mathbb{R}^{J+1} \to \mathbb{R}$. The matrix $\mathbf{J}_{\mathbf{G}}^{\mathbf{s}}(\bar{\mathbf{s}}) \in \mathbb{R}^{(J+1) \times (J+1)}$ with entries $(i + 1, j + 1)$ given by $\frac{\partial G_i(\bar{\mathbf{s}})}{\partial s_j}$ denotes the Jacobian matrix of $\mathbf{G}$ with respect to $\mathbf{s}$ at point $\bar{\mathbf{s}}$.

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

Preliminaries   This section provides some preliminary mathematical definitions and a result used in the proofs that follow.

Definition 1. $G_j : \mathbb{R}_{++}^{J+1} \to \mathbb{R}^{J+1}$ is linearly homogeneous if $G_j(\lambda \mathbf{s}) = \lambda G_j(\mathbf{s})$ for all $\lambda > 0$ and $\mathbf{s} \in \mathbb{R}_{++}^{J+1}$. $\mathbf{G}$ is homogeneous if each of its component G_j is.

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

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

Lemma 1 (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.

B Properties of the IPDL Model

Recall first that $d(j)$ is the set of products that are grouped with product j according to grouping characteristic d and that $s_{d(j)} = \sum_{k \in d(j)} s_k$ denotes the market share of group $d(j)$. To ease exposition, we omit the notation for parameters $\boldsymbol{\theta}_2$ and markets t. Recall then that the IPDL model is defined by

$$\sigma_j^{-1}\left(\mathbf{s}\right) = \ln G_j\left(\mathbf{s}\right) + c = \delta_j, \quad j \in \mathcal{J}, \tag{10}$$

where the function $\mathbf{G} : \mathbb{R}_{++}^{J+1} \to \mathbb{R}_{++}^{J+1}$ is defined by

$$\ln G_j\left(\mathbf{s}\right) = \left(1 - \sum_{d=1}^{D} \mu_d\right) \ln\left(s_j\right) + \sum_{d=1}^{D} \mu_d \ln\left(s_{d(j)}\right), \quad j = 1, \ldots, J, \tag{11}$$

$$\ln G_0\left(\mathbf{s}\right) = \ln\left(s_0\right), \tag{12}$$

with $\sum_{d=1}^{D} \mu_d < 1$ and $\mu_d \geq 0$, $d = 1, \ldots, D$.

Lemma 2. Let $\ln \mathbf{G} \equiv \left(\ln G_0, \ldots, \ln G_J\right)$.

1. The Jacobian matrix $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}\left(\mathbf{s}\right)$ of the function $\ln \mathbf{G}$ with respect to $\mathbf{s}$ has entries

$$\frac{\partial \ln G_i(\mathbf{s})}{\partial s_j} = \begin{cases} \dfrac{1 - \sum_{d=1}^{D} \mu_d}{s_i} + \sum_{d=1}^{D} \dfrac{\mu_d}{s_{d(i)}}, & i = j > 0, \\ \sum_{d=1}^{D} \dfrac{\mu_d}{s_{d(i)}} \mathbf{1}\{j \in d(i)\}, & i \neq j, \quad i > 0, j > 0, \\ \dfrac{1}{s_0}, & i = j = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

It is positive definite for all $\mathbf{s} \in \mathbb{R}_{++}^{J+1}$.

2. The function

$$\mathbf{s} \to -\mathbf{s}^{\mathsf{T}} \ln \mathbf{G}(\mathbf{s}) = -\sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s})$$

$$= -\left[\left(1 - \sum_{d=1}^{D} \mu_d \right) \sum_{j \in \mathcal{J}} s_j \ln(s_j) + \sum_{d=1}^{D} \mu_d \left(\sum_{g \in d} s_g \ln(s_g) \right) \right]$$

is strictly concave on Δ°.

3. The function $\mathbf{s} \to \mathbf{G}(\mathbf{s})$ is linearly homogeneous.

Proof of Lemma 2.

1. $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})$ is positive definite for all $\mathbf{s} \in \mathbb{R}_{++}^{J+1}$, as it is a symmetric, strictly diagonally dominant matrix with positive diagonal entries (Horn and Johnson, 2012, Theorem 6.1.10.)

2. Consider $\mathbf{s} \in \Delta^{\circ}$. The Hessian of $-\mathbf{s}^{\mathsf{T}} \ln \mathbf{G}(\mathbf{s})$ is $-\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})$, which is negative definite (by part 1).

3. Note that $G_0(\mathbf{s}) = s_0$ and

$$G_j(\mathbf{s}) = s_j^{1 - \sum_{d=1}^{D} \mu_d} \prod_{d=1}^{D} \left(s_{d(j)} \right)^{\mu_d}, \quad j = 1, \dots, J.$$

$\mathbf{G}$ linearly homogeneous since for any $\lambda > 0$, for all $j = 1, \dots, J$

$$G_j(\lambda \mathbf{s}) = (\lambda s_j)^{1 - \sum_{d=1}^{D} \mu_d} \prod_{d=1}^{D} \left(\sum_{k \in d(j)} \lambda s_k \right)^{\mu_d},$$

$$= \left[\lambda^{1 - \sum_{d=1}^{D} \mu_d} \prod_{d=1}^{D} \lambda^{\mu_d} \right] \left[s_j^{1 - \sum_{d=1}^{D} \mu_d} \prod_{d=1}^{D} \left(s_{d(j)} \right)^{\mu_d} \right]$$

$$= \lambda^{1 - \sum_{d=1}^{D} \mu_d + \sum_{d=1}^{D} \mu_d} G_j(\mathbf{s})$$

$$= \lambda G_j(\mathbf{s}),$$

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

B.1 Invertibility

As stated in the following proposition, the function is $\ln \mathbf{G}$ injective and hence invertible on its range.

Proposition 1. The function $\ln \mathbf{G}$ with $\mathbf{G}$ defined by Equations (11) and (12) is injective, with range equal to $\mathbb{R}^{J+1}$.

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

Also by Lemma 2, the function $\mathbf{s} \to \sum_{j \in \mathcal{J}} s_j \ln G_j(\mathbf{s})$ is strictly convex. Hence, for any $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$, the maximization problem $\sup_{\mathbf{s} \in \mathbb{R}^{J+1}} \{ \sum_{j \in \mathcal{J}} s_j (\delta_j - \ln G_j(\mathbf{s})) \}$ has a unique solution. Lastly, note that $|\ln \mathbf{G}(\mathbf{s})| \to \infty$ whenever $\mathbf{s} \to \mathbf{s}^0$, where $\mathbf{s}^0$ is on the boundary of $\mathbb{R}^{J+1}$, which means that at least one component of $\ln \mathbf{G}$ tends to infinity as $\mathbf{s}$ approaches the boundary of $\mathbb{R}^{J+1}$ and ensures that the solution is interior. Then, the solution is given by the first-order condition, which is $\boldsymbol{\delta} = \ln \mathbf{G}(\mathbf{s})$. $\qquad \square$

Invertibility of $\ln \mathbf{G}$ is equivalent to invertibility of the IPDL inverse market share function. Consider any vector of market shares $\mathbf{s} \in \Delta^{\circ}$. Then, holding $\delta_0 = 0$, the injectivity of the IPDL inverse market share function ensures that there exists a unique vector of indexes $\boldsymbol{\delta} \in \mathcal{D}$ that rationalizes demand, i.e., $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$.

Lemma 3. The IPDL inverse market share function $\boldsymbol{\sigma}^{-1}$ defined by Equations (10) – (12) is invertible.

As shown above, the IPDL model allows complementarity in demand, as defined by a negative cross-price derivative of market share. We cannot, therefore, use Berry (1994)'s and Berry et al. (2013)'s invertibility results, which strictly rule out such a form of complementarity. Berry et al. (2013) show invertibility for any market share function satisfying their "connected substitutes" conditions. The connected substitutes structure requires two conditions: (i) products are weak gross substitutes, that is, everything else equal, an increase in δ_i weakly decreases market share 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. The first requirement strictly rules out complementarity in demand defined by a negative cross-price derivative of market share.

Their result can, however, cover demand systems with some form of complementarity in demand: this happens when the demand function $q_j(\boldsymbol{\delta}_t)$ has a negative cross-price derivative with respect to price p_i, whereas its transformation to a market share function $q_j(\boldsymbol{\delta}_t)/\sum_{k\in\mathcal{J}} q_k(\boldsymbol{\delta}_t)$ does not (Example 1 in Berry et al. (2013) and Section 4.3 in Compiani (2022)). In contrast to us, Berry et al. (2013) do not require that the demand function is differentiable.

Let $\mathbf{H} = \mathbf{G}^{-1}$ denote the inverse of $\mathbf{G}$: $\mathbf{H}(e^{\boldsymbol{\delta}}) = (H_0(e^{\boldsymbol{\delta}}), \ldots, H_J(e^{\boldsymbol{\delta}})) = \mathbf{G}^{-1}(e^{\boldsymbol{\delta}})$. We show that $\mathbf{H}$ is linearly homogeneous.

Lemma 4. The function $e^{\boldsymbol{\delta}} \to \mathbf{H}(e^{\boldsymbol{\delta}})$ is linearly homogeneous.

Proof. Let $\mathbf{G}(\mathbf{s}) = e^{\boldsymbol{\delta}}$ or equivalently $\mathbf{s} = \mathbf{H}(e^{\boldsymbol{\delta}})$. Then, for any $\lambda > 0$,

$$\mathbf{H}(\lambda e^{\boldsymbol{\delta}}) = \mathbf{H}(\lambda \mathbf{G}(\mathbf{s})) = \mathbf{H}(\mathbf{G}(\lambda \mathbf{s})) = \lambda \mathbf{s} = \lambda \mathbf{H}(e^{\boldsymbol{\delta}}).$$

B.2 Micro-foundation

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 IPDL inverse market share function is consistent with a representative consumer who chooses a vector $\mathbf{s} \in \Delta$ of market shares of the differentiated products and a quantity $z \geq 0$ of the numéraire good so as to maximize her direct utility function

$$\alpha z + \sum_{j\in\mathcal{J}} v_j s_j - \left[\left(1 - \sum_{d=1}^{D} \mu_d \right) \sum_{j\in\mathcal{J}} s_j \ln(s_j) + \sum_{d=1}^{D} \mu_d \left(\sum_{g\in d\cup\{0\}} s_g \ln(s_g) \right) \right], \tag{14}$$

subject to the budget constraint and the constraint that the market share vector sums to one,

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

where $\alpha > 0$ is the marginal utility of income, $s_g = \sum_{k \in g} s_k$, and d is identified with the set of groups for grouping characteristic d.

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 her taste for variety (Lemma 2 above).

We further show that the direct utility function (14) gives the indirect utility function

$$\alpha y + \ln \left(\sum_{k \in \mathcal{J}} H_k \left(e^{\boldsymbol{\delta}} \right) \right), \tag{16}$$

where the second term is, up to an additive constant, the consumer surplus $CS(\boldsymbol{\delta})$.

We summarize these results as follows.

Proposition 2. The IPDL model (10) – (12) is consistent with a representative consumer who maximizes her direct utility (14) subject to constraints (15). Further, the direct utility (14) gives the indirect utility (16), where the second term is the convex consumer surplus function.

Proof. 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

$$u\left(\mathbf{s}\right) = \alpha y + \sum_{j \in \mathcal{J}} \delta_j s_j - \left[\left(1 - \sum_{d=1}^{D} \mu_d \right) \sum_{j \in \mathcal{J}} s_j \ln \left(s_j \right) + \sum_{d=1}^{D} \mu_d \left(\sum_{g \in d \cup \{0\}} s_g \ln \left(s_g \right) \right) \right]$$

41

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

$$\delta_j - \left[\left(1 - \sum_{d=1}^{D} \mu_d\right)(\ln(s_j) + 1) + \sum_{d=1}^{D} \mu_d \left(\ln\left(s_{d(j)}\right) + 1\right)\right] - \lambda = 0$$

which can be simplified as

$$\delta_j - \left[\left(1 - \sum_{d=1}^{D} \mu_d\right)\ln(s_j) + \sum_{d=1}^{D} \mu_d \ln\left(s_{d(j)}\right) + 1\right] - \lambda = 0,$$

for all $j = 1, \ldots, J$, and $\delta_0 - (\ln(s_0) + 1) - \lambda = 0$ for the outside good.

The first-order condition for an interior solution has a unique solution since the objective is strictly concave by Lemma 2; hence, the utility-maximizing demand exists uniquely. Setting $c = 1 + \lambda$, we show that the representative consumer model leads to the IPDL inverse market share function.

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

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

where the last equality uses the homogeneity of $\mathbf{H}$ (Lemma 4). Using that demands sum to 1, we find that

$$e^c = \sum_{k \in \mathcal{J}} H_k \left(e^{\boldsymbol{\delta}}\right), \tag{17}$$

so that the IPDL market share function is given by

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

Finally, substituting the market share functions (18) with the market shares s_j gives the indirect utility function (16).

By Roy's identity, the Hessian of the consumer surplus is $\mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}}(\boldsymbol{\delta})$, which by Proposition 3 (part 2.) is positive semi-definite. Convexity of the consumer surplus then follows. □

Anderson et al. (1988) and Verboven (1996b) show that the logit and nested logit models are consistent with a utility-maximizing representative consumer model. Proposition 2 extends these results to the IPDL model.

Furthermore, as shown by Allen and Rehbeck (2019b), utility (14) can be obtained, by aggregating across heterogeneous, utility-maximizing consumers, from the class of latent utility models with additively separable unobservable heterogeneity called perturbed utility.[15] This implies that the IPDL model embodies consumer heterogeneity and can be rationalized by a model with heterogeneous, utility-maximizing consumers.

B.3 Substitution Patterns

Proposition 3. The IPDL model has the following properties.

1. The independence from irrelevant alternatives (IIA) property holds for products of the same type; but does not hold in general for products of different types.

2. The matrix of price derivatives of market share $\mathbf{J}_{\boldsymbol{\sigma}}^{\mathrm{p}}(\boldsymbol{\delta})$ with entries $\partial\sigma_i(\boldsymbol{\delta})/\partial p_j$ is equal to

$$\mathbf{J}_{\boldsymbol{\sigma}}^{\mathrm{p}}(\boldsymbol{\delta}) = -\alpha\left([\mathbf{J}_{\ln\mathbf{G}}^{\mathbf{s}}(\mathbf{s})]^{-1} - \mathbf{s}\mathbf{s}^{\mathsf{T}}\right), \qquad (19)$$

with $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$ and where $\mathbf{J}_{\ln\mathbf{G}}^{\mathbf{s}}(\mathbf{s})$ has entries given by Equation (13). In the absence of income effects, the matrix of price derivatives of demand is the Slutsky matrix. It is symmetric and positive semi-definite, which implies that the IPDL market share functions are non-decreasing in their own index δ_j, $\partial\sigma_j(\boldsymbol{\delta})/\partial\delta_j \geq 0$. The cross-price elasticity from product j to product k is given by $(\partial\sigma_k(\boldsymbol{\delta}_t)/\partial p_{jt})(p_{jt}/\sigma_k(\boldsymbol{\delta}_t))$. The diversion ratio from product j to product k is given in percentage terms by $-100(\partial\sigma_k(\boldsymbol{\delta})/\partial p_j)/(\partial\sigma_j(\boldsymbol{\delta})/\partial p_j)$.

---

[15]See Hofbauer and Sandholm (2002), McFadden and Fosgerau (2012) and Fudenberg et al. (2015) for more details on perturbed utility models. Allen and Rehbeck (2019a) show that some perturbed utility models allow for complementarity in demand.

3. Products can be substitutes or complements in demand.

4. Products are substitutes in utility.

Proof of Proposition 3.

1. Using Equation (6), for any pair of products j and k we have

$$\frac{\sigma_j(\boldsymbol{\delta})}{\sigma_k(\boldsymbol{\delta})} = \exp\left(\frac{\delta_j - \delta_k}{1 - \sum_{d=1}^{D}\mu_d} + \sum_{d=1}^{D}\frac{\mu_d}{1 - \sum_{d=1}^{D}\mu_d}\ln\left(\frac{\sigma_{d(k)}(\boldsymbol{\delta})}{\sigma_{d(j)}(\boldsymbol{\delta})}\right)\right). \tag{20}$$

For products j and k of the same type (i.e., with $d(k) = d(j)$ for all d), Equation (20) 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, that is, 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. Recall that the IPDL model is defined by

$$\ln G_j(\mathbf{s}) + CS(\boldsymbol{\delta}) = \delta_j, \quad j \in \mathcal{J}, \tag{21}$$

where $\ln G_j$ is given by Equations (11) – (12) and where we have used that $c = \ln\left(\sum_{k\in\mathcal{J}}H_k\left(e^{\boldsymbol{\delta}}\right)\right) \equiv CS(\boldsymbol{\delta})$.

Differentiate this equation with respect to $\boldsymbol{\delta}$, then $\mathbf{I} = \mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})\mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}}(\boldsymbol{\delta}) + \mathbf{1}\mathbf{s}^{\mathsf{T}}$, with $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$, and where we have used Roy's identity. $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})$ is invertible. Then, $\mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}}(\boldsymbol{\delta}) = [\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})]^{-1}[\mathbf{I} - \mathbf{1}\mathbf{s}^{\mathsf{T}}] = [\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})]^{-1} - [\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})]^{-1}\mathbf{1}\mathbf{s}^{\mathsf{T}}$. Finally, note that $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})\mathbf{s} = \mathbf{1}$, so that $[\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})]^{-1}\mathbf{1}\mathbf{s}^{\mathsf{T}} = \mathbf{s}\mathbf{s}^{\mathsf{T}}$.

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

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

is symmetric and idempotent and thus positive semi-definite. Then $\mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}}(\boldsymbol{\delta})$ is positive semi-definite.

3. Suppose there are $J = 3$ products and one outside good. Products are grouped

according to two grouping characteristics: the grouping is $\{1\}, \{2, 3\}$ for the first characteristic and $\{1, 2\}, \{3\}$ for the second characteristic.

Let $\boldsymbol{\sigma}(\boldsymbol{\delta}) = \mathbf{s}$. Using Proposition 3 (part 2.), we show that

$$\frac{\partial \sigma_1 (\boldsymbol{\delta})}{\partial p_3} = s_1 s_3 \left[1 + \frac{\mu_1 \mu_2 s_2}{D} \right], \tag{22}$$

where $D = -(1 - \mu_1 - \mu_2)(s_1 + s_2)(s_2 + s_3) - \mu_1 \mu_2 s_2 (1 - s_0) < 0$. Products 1 and 3 are then complements in demand if and only if $\frac{\partial \sigma_1 (\boldsymbol{\delta})}{\partial p_3} < 0$, that is, if and only if $(1 - \mu_1 - \mu_2)(s_1 + s_2)(s_2 + s_3) - \mu_1 \mu_2 s_0 s_2 < 0$.

4. The IPDL model restricts products to be substitutes in utility, since

$$\frac{\partial u(\mathbf{s})}{\partial s_i \partial s_j} = -\sum_{d=1}^{D} \frac{\mu_d}{s_{d(j)}} \mathbf{1}\{i \in d(j)\} - \frac{\mu_0}{s_0}. \tag{23}$$

is negative. $\qquad \square$

B.4 Higher-Order Partial Derivatives of Demand in the RCL Model

Consider an additive random utility model (ARUM) in which the vector of random utility components follows a joint distribution with finite means that is absolutely continuous and independent of $\boldsymbol{\delta}$. The logit and nested logit models are special cases of this.

The ARUM market share functions (or choice probabilities) satisfy a range of general conditions, one of which is that the partial derivatives of the market share function σ_i ($i \in \mathcal{J}$) with respect to any set of distinct prices other than p_i are non-negative if the coefficients on prices are negative, i.e.,

$$\frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_2} \geq 0; \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_2 \partial p_3} \geq 0; \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_2 \partial p_3 \partial p_4} \geq 0, \text{ etc.} \tag{24}$$

In particular, this condition rules out complementarity in demand in the ARUM.

The RCL market share function is a mixture of logit market share functions, and thus it also satisfies the non-negativity condition when price coefficients are

45

negative (almost surely).

By contrast, the IPDL model does not necessarily satisfy the non-negativity condition. To show this, we consider, again, the example of Subsection 3.1 with $J = 3$ products and one outside good. Market shares are equal to $s_0 = 1/2$ and $s_1 = s_2 = s_3 = 1/6$ and $\alpha = 1$. Products are grouped according to two grouping characteristics: the grouping is $\{1\}, \{2, 3\}$ for the first characteristic and $\{1, 2\}, \{3\}$ for the second characteristic.

When $\mu_1 = 1/4$ and $\mu_2 = 1/3$, all products are substitutes in demand, but the non-negativity condition does not hold for the second-order partial derivatives:

$$\frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_3} = 0.0205 \geq 0; \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_3 \partial p_2} = -0.0204 \leq 0. \tag{25}$$

When $\mu_1 = 3/5$ and $\mu_2 = 1/3$, products 1 and 3 are complements in demand, but the non-negativity condition holds for the second order mixed derivatives:

$$\frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_3} = -0.0107 \leq 0; \frac{\partial \sigma_1(\boldsymbol{\delta})}{\partial p_3 \partial p_2} = 0.0052 \geq 0. \tag{26}$$

B.5 Direct and Indirect Substitution Effects

We follow Ogaki (1990) to decompose the substitution effect between products i and j into the indirect substitution effect and the direct substitution effect.

Let $\mathbf{J}_{\boldsymbol{\sigma}}^{\text{p}}$ be the matrix of market share derivatives with respect to prices. Recall that $\mathcal{J} = \{0, 1, \ldots, J\}$ denotes the choice set. Then, $\mathcal{J}_{-(i,j)}$ denotes the choice set without products i and j, and $\mathbf{J}_{\boldsymbol{\sigma}}^{\text{p}}[\mathcal{J}_{-(i,j)}, \mathcal{J}_{-(i,j)}]$ denotes the matrix $\mathbf{J}_{\boldsymbol{\sigma}}^{\text{p}}$ after removing the rows and columns involving products i and j.

The indirect substitution effect between product i and j, S_{ij}^{i}, is equal to

$$S_{ij}^{\text{i}} = \mathbf{J}_{\boldsymbol{\sigma}}^{\text{p}}[i, \mathcal{J}_{-(i,j)}] \left[\mathbf{J}_{\boldsymbol{\sigma}}^{\text{p}}[\mathcal{J}_{-(i,j)}, \mathcal{J}_{-(i,j)}] \right]^{-1} \mathbf{J}_{\boldsymbol{\sigma}}^{\text{p}}[\mathcal{J}_{-(i,j)}, j]. \tag{27}$$

The direct substitution effect between product i and j, S_{ij}^{d}, is then equal to

$$S_{ij}^{\text{d}} = \frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial p_i} - S_{ij}^{\text{i}}. \tag{28}$$

Products i and j are direct complements if $S_{ij}^{\mathrm{d}} < 0$, and direct substitutes if $S_{ij}^{\mathrm{d}} > 0$.

C Details on the Experiments

C.1 Simulated Data

For each experiment, we generate 50 datasets consisting of $T = 200$ independent markets with $J = 45$ products, where the markets exhibit product segmentation according to two grouping characteristics forming four product types. The grouping structure is simulated using binomial distributions and is common across markets.

In each experiment, we simulate a fully structural model of demand and supply, where the observed characteristic x_{jt} and the cost-shifter z_{jt} are i.i.d. $\mathcal{U}(0,1)$. The unobserved product characteristic is $\xi_{jt} = u_{1t} + u_{2t}$ and the unobserved cost component is $\omega_{jt} = u_{1t} + u_{3t}$, where u_{1t}, u_{2t}, and u_{3t} are i.i.d. $\mathcal{U}(-0.5, 0.5)$. Prices and market shares are determined endogenously. The supply side is a static price competition model among five multi-product firms, each with nine products and with constant marginal cost given by $c_{jt} = 2 + x_{jt} + z_{jt} + w_{jt}$.

Experiment 1 We simulate four IPDL models, varying the values of the grouping parameters μ_1 and μ_2 (Table 1). In each model, we set $\delta_{jt} = -3 + 2x_{jt} - 0.5p_{jt} + \xi_{jt}$.

Experiment 2 We simulate four PDL models, varying the values of the grouping parameters μ_1 and μ_2 (Table 2). The PDL model is a GEV model. Its market share function is given by $\sigma_j(\boldsymbol{\delta}) = e^{\delta_j}(\partial G_j(e^{\boldsymbol{\delta}})/\partial e^{\delta_j})/G(e^{\boldsymbol{\delta}})$, where $G(e^{\boldsymbol{\delta}}) = a_1 \left[\sum_{g=1}^{2} \sum_{j \in G_{1g}} \left(e^{\delta_j/\mu_1} \right)^{\mu_1} \right] + a_2 \left[\sum_{g=1}^{2} \sum_{j \in G_{2g}} \left(e^{\delta_j/\mu_2} \right)^{\mu_2} \right]$, with $a_1 = (1-\mu_1)/(2-\mu_1-\mu_2)$ and $a_2 = 1 - a_1$. In each model, we set $\delta_{jt} = -1 + 2x_{jt} - 0.5p_{jt} + \xi_{jt}$.

Experiment 3 We simulate four RCL models with independent normal random coefficients on dummies for groups, varying the values of the standard deviations of the normal random coefficients RC_1 and RC_2 (Table 3). In each model, the mean utility of product j in market t is given by $\delta_{jt} = 3 - p_{jt} + x_{jt} + \xi_{jt}$. We

use the package PyBLP from Conlon and Gortmaker (2020) to simulate the RCL models. We use 100 Halton draws over the standard normal distribution to integrate the market share functions numerically. Each dimension of integration of Halton draws uses a different prime, discards the first 1,000 points, and then scrambles the sequence.

C.2 Estimation Procedures

Nested Logit and IPDL Models   We estimate the nested logit and IPDL models using the two-stage least squares estimator. We compute instruments as the predicted values from regressions of the endogenous variables (i.e., price variable and two log-share terms) on a constant, the product characteristic x_{jt}, the cost shifter z_{jt}, and the differentiation IVs $\sum_{k \in d_1(j)} \left(d^x_{j,k,t} \right)^2$, $\sum_{k \notin d_1(j)} \left(d^x_{j,k,t} \right)^2$, and $\sum_{k \in d_2(j)} \left(d^x_{j,k,t} \right)^2$, with $d^x_{j,k,t} \equiv x_{kt} - x_{jt}$.

PDL Models   We estimate the PDL models using the following two-step estimation procedure, which is standard in the literature. First, we solve for the error term ξ_{jt} as a function of the parameters μ_1 and μ_2. That is, given values for μ_1 and μ_2, we numerically compute the values $\delta_{jt}(\mu_1, \mu_2)$ of δ_{jt} that equate the observed to the predicted market shares, and compute $\hat{\xi}_{jt} = \delta_{jt}(\mu_1, \mu_2) - \beta_0 - \hat{\beta} x_{jt} + \hat{\alpha} p_{jt}$, where $\hat{\alpha}$ and $\hat{\beta}$ are the estimates of the Berry (1994)'s regression based on $\delta_{jt} = \beta_0 + \beta x_{jt} - \alpha p_{jt} + \xi_{jt}$. Second, we interact $\hat{\xi}_{jt}$ with instruments to form the generalized method of moments objective function that we minimize over μ_1 and μ_2.

RCL Models   We estimate the RCL models using the nested-fixed point approach proposed by BLP. We use the package PyBLP by Conlon and Gortmaker (2020) and implement their best practices. That is, we numerically integrate the market share functions using 100 Halton draws over the standard normal distribution; we numerically compute the δ_{jt}'s that equate the observed to the predicted market shares using the SQUAREM accelerated fixed point algorithm; we minimize the generalized method of moments objective function using the Knitro 13.1 Interior/Direct algorithm; and we use the "approximate" version of the feasible optimal instruments.

D Details on the Empirical Application

Potential market size. We compute the potential market size by multiplying the total number of persons in a market by the monthly per capita consumption of cereals. For each store in a month, we compute the total number of persons as the weekly average number of households who visited that store in that given month, multiplied by the average household size. First, we compute the weekly average number of households using the information on the daily traffic store and assuming that consumers visit stores twice a week. Lastly, we compute the monthly per capita consumption of cereals using the information from the USDA's Economic Research Service that per capita US consumption of cereals was equal to 13.4 pounds in 1991, 13.9 in 1992, 14.6 in 1993, 14.8 in 1994, 14.6 in 1995 and 14.3 in 1996.

Estimations We estimate the IPDL and RCL models using the procedures described in Appendix C.2. For the RCL models, we also absorb the fixed effects using the package PyHDFE.

Computational Time We run all the estimations on a personal computer with 2.50 GHz Intel Core i7-11850H CPU and 16.0 GB RAM, Windows operating system (Windows 10), Python version 3.9. We obtain the computational time using Python's timeit() function that returns the number of seconds it took to execute the code.

Goodness-of-fit We measure goodness-of-fit using a two-fold cross-validated procedure: (i) we randomly split the sample into two sub-samples of (approximate) equal size; (ii) we estimate both specifications of the IPDL and RCL models using the first sub-sample and compute the Mean Square Error (MSE) for the second sub-sample; (iii) we repeat (ii) by inverting the role of the two sub-samples; (iv) we compute the average of the two MSE. The model that obtains the lowest average MSE, called cross-validated MSE, best fits the data.

Complementarity in Demand We compute the percentage of pairs of complements in demand while accounting for the fact that some pairs of cereals exhibiting

negative diversion ratios may be independent and not complements in demand. For this purpose, we use 95% confidence intervals for the diversion ratios. We compute them using a parametric bootstrap: we repeatedly draw from the estimated joint distribution of parameters; for each draw, we compute the average (over markets) diversion ratios for all pairs of products, thus generating a bootstrap distribution. We take 500 draws.

References

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

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

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.

——— (1992): *Discrete choice theory of product differentiation*, Cambridge, MA: MIT press.

BARNETT, W. A. AND A. SERLETIS (2008): "Consumer preferences and demand systems," *Journal of Econometrics*, 147, 210–224.

BEN-AKIVA, M. AND M. BIERLAIRE (1999): "Discrete choice methods and their applications to short term travel decisions," in *Handbook of transportation science*, Springer, 5–33.

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., A. KHWAJA, V. KUMAR, A. MUSALEM, K. C. WILBUR, G. AL-LENBY, B. ANAND, P. CHINTAGUNTA, W. M. HANEMANN, P. JEZIORSKI, ET AL. (2014): "Structural models of complementary choices," *Marketing Letters*, 25, 245–256.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

BERRY, S. T. (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.

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

BLUNDELL, R., J. L. HOROWITZ, AND M. PAREY (2012): "Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation," *Quantitative Economics*, 3, 29–51.

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.

CARDELL, N. S. (1997): "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity," *Econometric Theory*, 185–213.

COMPIANI, G. (2022): "Market counterfactuals and the specification of multiproduct demand: A nonparametric approach," *Quantitative Economics*, 13, 545–591.

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

CONLON, C. AND J. H. MORTIMER (2021): "Empirical properties of diversion ratios," *The RAND Journal of Economics*, 52, 693–726.

CONLON, C. T. AND J. H. MORTIMER (2013): "Demand estimation under incomplete product availability," *American Economic Journal: Microeconomics*, 5, 1–30.

CONLON, C. T., M. SINKINSON, ET AL. (2021): "Common Ownership and Competition in the Ready-to-Eat Cereal Industry," Working paper.

CORTS, K. (1996): "The Ready-to-eat Breakfast Cereal Industry in 1994 (A)," *Harvard Business School*.

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.

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, M. MARCOUX, AND S. ORR (2021): "Estimating complementarity with large choice sets: An application to mergers," *Available at SSRN 3802097*.

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

FOSGERAU, M., J. MONARDO, AND A. DE PALMA (2021): "The inverse product differentiation logit model," *Available at SSRN 3141041*.

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. (2021): "On the representation of the nested logit model," *Econometric Theory*, 1–11.

GANDHI, A. AND J.-F. HOUDE (2021): "Measuring substitution patterns in differentiated products industries," *Working paper*.

GANDHI, A. AND A. NEVO (2021): "Empirical models of demand and supply in differentiated products industries," in *Handbook of Industrial Organization*, Elsevier, vol. 4, 63–139.

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

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

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

HAAG, B. R., S. HODERLEIN, AND K. PENDAKUR (2009): "Testing and imposing Slutsky symmetry in nonparametric demand systems," *Journal of Econometrics*, 153, 33–50.

HAUSMAN, J., G. LEONARD, AND J. D. ZONA (1994): "Competitive analysis with differenciated products," *Annales d'Economie et de Statistique*, No. 34, 159–180.

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

HORN, R. A. AND C. R. JOHNSON (2012): *Matrix analysis*, Cambridge university press.

HORTAÇSU, A., J. LIEBER, J. MONARDO, AND A. DE PAULA (2020): "Estimating Nesting Structures," *Working paper*.

IARIA, A. AND A. WANG (2020): "Identification and estimation of demand for bundles," *CEPR Discussion Paper No. DP14363, Available at SSRN 3535448*.

——— (2021): "An empirical model of quantity discounts with large choice sets," *CEPR Discussion Paper No. DP16666, Available at SSRN 3960270*.

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.

LEWBEL, A. AND L. NESHEIM (2019): "Sparse demand systems: corners and complements," Tech. rep., CeMMAP working papers CWP45/19.

LEWBEL, A. AND K. PENDAKUR (2009): "Tricks with Hicks: The EASI demand system," *American Economic Review*, 99, 827–63.

LU, Z., X. SHI, AND J. TAO (2022): "Semi-Nonparametric Estimation of Random Coefficient Logit Model for Aggregate Demand," *Journal of Econometrics, Forthcoming, Available at SSRN 3503560*.

MCFADDEN, D. (1978): "Modelling the choice of residential location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlquist, F. Snickars, and J. W. Weibull, Amsterdam: North Holland, vol. 673, 75–96.

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

MICHEL, C., J. M. PAZ Y MINO, AND S. WEIERGRAEBER (2022): "Estimating Industry Conduct Using Promotion Data - The Evolution of Pricing Behavior in the U.S. RTE Cereal Industry," *Wroking paper*.

MILLER, N. H. AND M. C. WEINBERG (2017): "Understanding the price effects of the MillerCoors joint venture," *Econometrica*, 85, 1763–1791.

MONARDO, J. (2021): "Measuring substitution patterns with a flexible demand model," *Available at SSRN 3921601*.

NEVO, A. (2000): "Mergers with differentiated products: The case of the ready-to-eat cereal industry," *The RAND Journal of Economics*, 395–421.

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

——— (2011): "Empirical models of consumer behavior," *Annual Review of Economics*, 3, 51–75.

OGAKI, M. (1990): "The indirect and direct substitution effects," *The American Economic Review*, 80, 1271–1275.

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

SALANIÉ, B. AND F. A. WOLAK (2022): "Fast," robust", and approximately correct: estimating mixed demand systems," *Working paper*.

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.

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

THOMASSEN, Ø., H. SMITH, S. SEILER, AND P. SCHIRALDI (2017): "Multi-category competition and market power: a model of supermarket pricing," *American Economic Review*, 107, 2308–51.

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.

VOVSHA, P. (1997): "Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area," *Transportation Research Record*, 1607, 6–15.

WANG, A. (2022): "Sieve BLP: A semi-nonparametric model of demand for differentiated products," *Journal of Econometrics*.

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

Supplement to "The Inverse Product Differentiation Logit Model" - For Online Publication

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

November 29, 2022

Abstract

We present simulations investigating some properties of the Inverse Product Differentiation Logit (IPDL) model.

Notation   We use italics for scalar variables and real-valued functions, boldface for vectors, matrices and vector-valued functions, and calligraphic for sets. $\mathbb{R}_+$ is the set of non-negative real numbers, $\mathbb{R}_{++}$ is the set of positive real numbers, and $\mathbb{R}_{++}^{J+1} = (0, \infty)^{J+1}$. As default, vectors are column vectors: $\mathbf{s} = (s_0, \ldots, s_J)^{\mathsf{T}} \in \mathbb{R}^{J+1}$.

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

Let $\mathbf{G} = (G_0, \ldots, G_J) : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a vector function composed of functions $G_j : \mathbb{R}^{J+1} \to \mathbb{R}$. The matrix $\mathbf{J}_{\mathbf{G}}^{\mathbf{s}}(\bar{\mathbf{s}}) \in \mathbb{R}^{(J+1) \times (J+1)}$ with entries $(i + 1, j + 1)$ given by $\frac{\partial G_i(\bar{\mathbf{s}})}{\partial s_j}$ denotes the Jacobian of $\mathbf{G}$ with respect to $\mathbf{s}$ at point $\bar{\mathbf{s}}$.

---

[*]University of Copenhagen; mogens.fosgerau@econ.ku.dk.

[†]University of Bristol; julien.monardo@bristol.ac.uk.

[‡]CY Cergy Paris Université; andre.de-palma@cyu.fr.

A univariate function $\mathbb{R} \to \mathbb{R}$ applied to a vector is a coordinate-wise application of the function, e.g., $\ln(\mathbf{s}) = (\ln(s_0), \ldots, \ln(s_J))$. $|\tilde{\mathbf{s}}| = \sum_{j \in \mathcal{J}} |\tilde{s}_j|$ denotes the 1-norm of vector $\tilde{\mathbf{s}}$.

The IPDL model Recall first that $d(j)$ is the set of products that are grouped with product j according to grouping characteristic d and that $s_{d(j)} = \sum_{k \in d(j)} s_k$ denotes the market share of group $d(j)$. To ease exposition, we omit notation for parameters $\boldsymbol{\theta}_2$ and markets t.

In the IPDL model, the matrix of derivatives of the market share function $\boldsymbol{\sigma}$ with respect to prices $\mathbf{p}$, $\mathbf{J}_{\boldsymbol{\sigma}}^{\mathbf{p}}(\boldsymbol{\delta})$, has entries $\partial \sigma_i(\boldsymbol{\delta})/\partial p_j$ equal to

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

with $\mathbf{s} = \boldsymbol{\sigma}(\boldsymbol{\delta})$ and where $\mathbf{J}_{\ln \mathbf{G}}^{\mathbf{s}}(\mathbf{s})$ has entries given by

$$\frac{\partial \ln G_i(\mathbf{s})}{\partial s_j} = \begin{cases} \dfrac{1 - \sum_{d=1}^{D} \mu_d}{s_i} + \sum_{d=1}^{D} \dfrac{\mu_d}{s_{d(i)}}, & i = j > 0, \\[2ex] \sum_{d=1}^{D} \dfrac{\mu_d}{s_{d(i)}} \mathbf{1}\{j \in d(i)\}, & i \neq j, \quad i > 0, j > 0, \\[2ex] \dfrac{1}{s_0}, & i = j = 0 \\[2ex] 0, & \text{otherwise.} \end{cases} \tag{2}$$

We cannot obtain closed-form formulae for the entries of the matrix price derivatives, and in turn, for the diversion ratios between products. We therefore perform simulations to better understand the substitution patterns of the IPDL model. We focus on the diversion ratios.

The diversion ratio from product j to product k is the fraction of consumers leaving product j following a price increase of product j, who switch to product k. It is given in percentage terms by $-100(\partial \sigma_k(\boldsymbol{\delta}_t)/\partial p_{jt})/(\partial \sigma_j(\boldsymbol{\delta}_t)/\partial p_{jt})$.

Simulated Data We simulate markets with 45 products and an outside good. For this, we first simulate:

- 20 different grouping structures according to 3 grouping characteristics, and

with 3 groups per characteristic. 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 $\boldsymbol{\mu} \in \Delta_3^\circ$. This normalization ensures that we simulate markets with very low and very high values for μ_0.

- 20 different vectors of market shares $\mathbf{s} = (s_0, \ldots, s_{45})$. We obtain a vector of market shares by simulating a 461-vector of uniformly distributed random numbers, where the first element is s_0, then by normalizing the vector of market shares of products so that $\mathbf{s} \in \Delta_{45}^\circ$. This normalization ensures that we simulate markets with very low and very high values for s_0.

Then, we combine the grouping structures, the grouping parameters and the market shares to form $8,000$ markets. Table 1 gives summary statistics.

Table 1: Summary Statistics on the Simulated Data

Variable	Mean	Min	Max
q_0	0.5297	0.0015	0.9452
q	0.0105	9e-06	0.0494
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 Structures Table 2 shows the distribution of the diversion ratios between products according to the number of common groups.

Diversion ratios can be either negative (complementarity) or positive (substitutability). Products of the same type are always substitutes. Otherwise, products can be either substitutes or complements. Products are more likely to be complements as they become more different.

3

Table 2: Diversion Ratios according to the Number of Common Groups

# Common groups	Median	Mean	Complements
0 (None)	-0.020	-0.320	52.29%
1	0.953	1.442	8.93%
2	2.314	3.269	0.00%
3 (All)	3.665	5.219	0.00%

Notes: Column "Complements" gives the percentage of negative diversion ratios according to the number of common groups. E.g., 52.29% of the pairs of products sharing zero group are complements.

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

As the parameter μ_{jk} becomes larger, we observe that (i) the diversion ratios increase in values, and that (ii) the share of complements decreases. This is because higher μ_d means that products of the same group according to grouping characteristic d become more similar.

Table 3: Percentage of Complements according to the Value of μ_{jk}

μ_{jk}	Median	Mean	Complements
$[0, 0.1[$	0.181	-4.529	36.09%
$[0.1, 0.2[$	1.262	1.540	3.14%
$[0.2, 0.3[$	1.611	1.973	5.66%
$[0.3, 0.4[$	1.955	2.531	4.28%
$[0.4, 0.5[$	2.192	2.856	5.64%
$[0.5, 0.6[$	3.902	4.692	0.00%
$[0.6, 0.7[$	5.110	5.759	0.00%
$[0.7, 0.8[$	5.793	6.478	0.00%
$[0.8, 0.9[$	4.915	5.856	0.00%
$[0.9, 1[$	10.559	11.322	0.00%

Notes: Column "Complements" gives the percentage of negative diversion ratios according to the number of common groups. E.g., 52.29% of the pairs of products sharing zero group are complements.

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 diversion ratios depends on the degree of closeness. The closer two products are, the higher is their diversion ratio.