Measuring Substitution Patterns with a Flexible Demand Model*

Julien Monardo†

September 17, 2021

Abstract

Measuring substitution patterns across differentiated products is at the heart of many empirical studies. Most of the approaches used in applied work, including the leading approach pioneered by Berry, Levinsohn, and Pakes (1995) (BLP), impose distributional and functional form assumptions and restrictions on individual consumer behavior that may restrict substitution patterns in the space of product characteristics. In this paper, I propose a flexible approach that does not place arbitrary assumptions about how substitution patterns depend on product characteristics and is agnostic about individual consumer behavior. To this end, I rely on an inverse market share function that I show to be consistent with utility maximization by heterogeneous consumers. I find my approach yields substitution patterns that BLP cannot recover, including complementarity in demand, and can accommodate multiple choices. My approach can be applied to topics in various fields of economics, such as digital economics, industrial organization, and international trade, to address policy-relevant questions such as the evaluation of mergers and regulatory changes in taxes and trade policy.

*First Draft: January, 2019. I would like to thank Steven Berry, Debopam Bhattacharya, Philippe Choné, Olivier De Groote, André de Palma, Xavier D'Haultfoeuille, Roxana Fernandez, Alfred Galichon, Gautam Gowrisankaran, Laura Grigolon, Ulrich Kaiser, Laurent Linnemer, Hugo Molina, Michelangelo Rossi, Pasquale Schiraldi, Yutec Sun and Thibaud Vergé. I also thank seminar participants at CREST, Telecom Paris, TSE, INRAE and Chicago, and workshop/conference participants at Workshop on Platforms, E-commerce and Digital Economics (2019, CREST-Telecom Paris), Conference on Industrial Organization and the Food Industry (2019, TSE), EARIE 2019 and 2021, ES World Congress 2020, DSE Winter School 2020, ASSA 2021, RES 2021, ES Asian Meeting 2021. This paper is based on Chapter 2 of my doctoral dissertation at University of Paris-Saclay and was previously circulated under the title "The Flexible Inverse Logit Model".

†Telecom Paris, Institut Polytechnique de Paris; jmonardo@telecom-paris.fr

1 Introduction

Understanding how consumers substitute across differentiated products is at the heart of many empirical studies, meaning that measuring substitution patterns is a preliminary step of these studies.[1] Prominent examples include the measurement of market power and the evaluation of mergers and acquisitions. Market power depends on substitution patterns as products facing closer substitutes tend to have lower markups. The incentives of the merging firms to increase their prices also depend on substitution patterns: the price increase following the merger will be higher the more substitutable the products of the merging firms are. To obtain convincing empirical findings, it is thus crucial to rely on credible estimates of substitution patterns.

The state-of-the-art approach to estimate substitution patterns among differentiated products using market-level data is due to Berry et al. (1995) (BLP). It uses the random coefficient logit model, which, theoretically, allows for a wide variety of substitution patterns determined by the proximity of products in characteristic space. In practice, however, achieving flexibility is challenging as it requires the utility and random coefficient specifications to be flexible enough. For tractability, many papers applying BLP impose arbitrary distributional assumptions (typically normal random coefficients) and functional form assumptions (usually linearity of the utility in characteristics), thereby possibly restricting the substitution patterns.[2] Besides, by using the random coefficient logit model, BLP assumes that each individual consumer chooses only her preferred product, which may bias substitution patterns when consumers choose baskets of products.[3]

In this paper, I propose a flexible approach that does not rely on arbitrary assumptions

[1]Examples span a wide range of topics in industrial organization: evaluation of mergers (Baker and Bresnahan, 1985; Nevo, 2000a; Miller and Weinberg, 2017), measurement of market power (Bresnahan, 1989; Berry et al., 1995; Nevo, 2001), and welfare gains from new products (Petrin, 2002). Examples can also be found in various fields of economics: environmental policy (Reynaert, 2021), financial markets (Dick, 2008; Ho and Ishii, 2011), food policy (Griffith et al., 2019), health care (Ho and Lee, 2017), telecommunications market (Bourreau et al., 2021), vertical markets (Berto Villas-Boas, 2007; Crawford and Yurukoglu, 2012), voting (Gordon and Hartmann, 2013), and trade policy (Verboven, 1996a; Berry et al., 1999).

[2]McFadden and Train (2000) show that the market share function (choice probabilities) implied by any random utility model can be approximated by a random coefficient logit model, provided that the utility and random coefficient specifications are flexible enough. In Section 5, I provide simulations showing that misspecification of random coefficients may bias estimates of substitution patterns. See also Lu et al. (2019), Compiani (2021), and Narita and Saito (2021). Davis and Schiraldi (2014) show that imposing functional form assumptions on the utility, such as linearity in characteristics, may generate substitution patterns that are not driven by how close products are in characteristic space, even when many random coefficients are included.

[3]See Narita and Saito (2021) for further details.

about how substitution patterns depend on product characteristics and is agnostic about individual consumer behavior. Furthermore, in line with BLP, I use market-level data where consumers choose among a set of products differentiated according to some characteristics that the modeler may observe or not. The modeling of unobserved characteristics possibly creates an endogeneity problem that I address using a standard instrumental variable strategy based on exogenous variation in the choice set. Thereby, my approach allows for a wide range of substitution patterns in characteristic space. In particular, I show that it yields substitution patterns that BLP cannot recover.

I obtain rich substitution patterns in two steps. First, I specify an inverse market share function that is flexible in the sense that it can match the substitution patterns generated by any member of a large class of inverse market share functions.[4] I refer to this inverse market share function as the flexible inverse logit (FIL) model since it improves the logit model thanks to a flexible nesting structure (i.e., allocation of products into groups). Specifically, this latter admits a group for each pair of products whose substitution pattern is governed by a nesting parameter that tends to be higher as the two products are closer substitutes.[5] By focusing on an inverse market share function, I build on a key insight underlying the literature: the inverse market share function is the target of estimation. It means that knowledge of it and its first and second derivatives is sufficient to answer many questions of interest. Based on Berry (1994), it also means that it generates equations that can be directly used as a basis for estimation using regressions. Besides, by doing so, I avoid placing restrictions on individual consumer behavior. The FIL model thus accommodates a wide variety of consumer behavior and substitution patterns, including multiple choices (i.e., purchases of several units of different products) and complementarity in demand (i.e., a negative cross-price derivative of market share). However, without further assumptions, this prevents any study at the individual consumer level that, e.g., requires knowledge of the distribution of consumer preferences. Lastly, the FIL model is consistent with utility maximization. In particular, it can be derived from a specific instance of the class of models of heterogeneous, utility-maximizing consumers studied by Allen and Rehbeck (2019). Thus, combined with a model of firm conduct, it can be used for counterfactual analysis such as the simulation of mergers or regulatory changes in taxes and trade policy.

Second, I map the nesting parameters in characteristic space. That is, I specify each

[4]Formally, this inverse market share function will be shown to be flexible in the sense of Diewert (1974).

[5]The FIL model can be motivated as a flexible member of the class of inverse market share functions developed by Fosgerau et al. (2021), which extend that of the nested logit models by allowing any nesting structure.

pair-specific nesting parameter as a parametric, pair-invariant function of how close the two products are in this space. This mapping allows generating substitution patterns determined by the proximity of products in characteristic space – rather than their identity, i.e., substitution patterns are in characteristic space – rather than in product space. As shown by Gandhi and Houde (2020), BLP also embeds these realistic restrictions. However, in BLP, they are determined by random coefficient specifications. Besides, consistency with economic theory, such as utility maximization, generates inequality restrictions on the mapping. Relying on the literature on the econometrics of shape restrictions (Matzkin, 1994; Chetverikov et al., 2018), I impose these restrictions during estimation to obtain better estimates of substitution patterns. Lastly, I specify the mapping using Bernstein polynomials to make the substitution patterns mainly driven by data rather than by distributional or functional forms assumptions. This also makes it convenient to impose the inequality restrictions since they can be readily converted into a set of constraints that are linear in the Bernstein polynomial coefficients.

My general finding is that my approach allows for a large set of possible substitution patterns. In theory, it and BLP generate two non-nested sets: in constrast to BLP, I rule out income effects and allow for complementarity in demand. As a by-product, using a standard result in convex analysis, I establish a new invertibility result that applies to a large class of (inverse) market share functions, including some that Berry et al. (2013)'s results do not cover.[6] However, absent income effects and complementarity in demand, it is not clear which of the two approaches provides a broader set of substitution patterns. To gain insights into this, I run Monte-Carlo simulations and, at the same time, examine implications for implied markups and merger price effects based on static price competition between multi-product firms. Simulations lead to the following results. First, even when restricted to substitutes, my approach generates substitution patterns different from BLP. It means that it can accommodate substitution patterns that BLP cannot recover. Furthermore, my approach can outperform BLP in two cases: first, when BLP is misspecified about its random coefficient specification, and, second, when consumers choose baskets of products, rather than products alone. These results suggest that my approach can obtain in some cir-

[6]This class of inverse market share functions includes those implied by many additive random utility models (ARUM) used in applied work. In the ARUM, the utility that a consumer derives from choosing a product is given by the sum of a deterministic utility term and a random utility term (Anderson et al., 1992). This invertibility result includes ARUM without income effect and heterogeneity in preferences apart from the random utility terms. In particular, it rules out observed heterogeneity related to observed individual characteristics and unobserved heterogeneity through random coefficients.

cumstances better estimates than BLP, which can be explained by the fact that it does not rely on arbitrary assumptions about how substitution patterns depend on product characteristics and is agnostic about individual consumer behavior. Lastly, my approach can match the substitution patterns implied by the benchmark BLP allowing a normal random coefficient on an observed product characteristic, even when there is substantial unobserved heterogeneity in preferences.

Ongoing Work. To illustrate my approach, I revisit work by Nevo (2000a, 2001) on market power and merger simulation in the ready-to-eat cereals market.

Related Literature. My paper relates to several strands of literature on the estimation of demand models for differentiated products. First, as already mentioned, it is in line with the literature pioneered by Berry (1994) and BLP. However, it differs from this literature by relying on a different modeling strategy. BLP assumes an additive random utility model – the random coefficient logit model – from which the market share function is derived by utility maximization and inverted numerically to obtain inverse market share equations as a basis for estimation. Since the inversion is not in closed form, BLP prevents us from using regressions for estimation and thus requires a more complex estimation procedure.[7] By contrast, I specify an inverse market share function that I show to be consistent with utility maximization and directly estimate using parametric regression techniques. In this respect, closest to my paper are Compiani (2021) and Fosgerau et al. (2021) who directly estimate inverse market share functions using regression techniques based on market-level data. Compiani (2021) develops a non-parametric approach. He also does not make distributional assumptions and imposes minimal functional form restrictions grounded in economic theory. However, his approach turns out to be feasible only in settings with small choice sets, whereas my approach can handle small to very large choice sets. Fosgerau et al. (2021) develops a class of inverse market share functions that extend that of the nested logit models by allowing any nesting structure while preserving their computational simplicity. The FIL model is the member that admits a flexible nesting structure with a group for each pair of products.

[7] See Conlon and Gortmaker (2020) for more details. BLP complicates the estimation, which involves a numerical inversion nested into a non-linear, non-convex optimization problem and simulation of the market share function. It thus requires dealing with the related issues of local optima, choice of starting values, and accuracy of simulation and numerical inversion (see Knittel and Metaxoglou, 2014, and references therein). Other approaches to solve BLP's problem were proposed by Dubé et al. (2012); Lee and Seo (2015); Salanié and Wolak (2019).

Therefore, my paper is also linked to the literature that estimates substitution patterns using models based on a nesting structure that allocates products into groups, where products in the same groups tend to be closer substitutes. The most prominent examples are the nested logit models that group products into exhaustive, mutually exclusive groups. They are easy to estimate by linear regression but have been criticized for restricting substitution patterns. This has led the literature to propose models based on more general nesting structures. The vast majority of them are members of the class of generalized extreme value (GEV) models developed by McFadden (1978). These models include the ordered logit (Small, 1987), the product differentiation logit (Bresnahan et al., 1997), the ordered nested logit (Grigolon, 2021), etc. Closer to the FIL model are the paired combinatorial logit (Chu, 1989; Koppelman and Wen, 2000) and the flexible coefficient multinomial logit (Davis and Schiraldi, 2014), which also use a flexible nesting structure. However, they restrict products to be substitutes in demand and require a complex estimation procedure similar to that in BLP.

Furthermore, my paper is related to the flexible functional form approach (see Barnett and Serletis, 2008, and references therein), which, e.g., includes the almost ideal demand system of Deaton and Muellbauer (1980). This line of research builds flexible models in the sense of Diewert (1974) to derive demand equations that only comprise observables and require the addition of additive error terms to serve as a basis for estimation. As highlighted by the literature, this approach has two main limitations. First, it has many parameters that quickly increases with the number of products, which makes its use to large choice sets very challenging. This is because it requires very large datasets and a high number of instruments. Second, because it generates substitution patterns in product space, it cannot be used to predict the demand for a new product. My approach also involves a flexible model of consumer behavior – the FIL model – but maps parameters in characteristic space, which overcomes these limitations.

Lastly, my paper relates to papers that aim to obtain rich substitution patterns in characteristics space. They include papers that semi- or non-parametrically estimate random coefficients in demand models, i.e., without imposing distributional assumptions on the random coefficients (Fox et al., 2011, 2016; Fox and Gandhi, 2016; Lu et al., 2019). They also include Athey and Imbens (2007) who extend BLP by allowing multiple unobserved characteristics terms, and Bajari and Benkard (2005) and Berry and Pakes (2007) who use pure characteristics demand models. Furthermore, they also include papers that map demand parameters in characteristics space, and in particular those that apply the distancemetric approach of Pinkse et al. (2002) for demand estimation purposes (Pinkse and Slade, 2004; Slade, 2004; Davis and Schiraldi, 2014).

The remainder of the paper is organized as follows. Section 2 provides an overview of my approach. Section 3 provides an extended discussion about the theoretical aspects of my approach, including a study of the substitution patterns that it accommodates and its consistency with standard consumer theory. Section 4 describes the empirical strategy. Section 5 presents simulations that show the performance of my approach and how it compares to BLP. Section 6 concludes.

2 Overview of the Approach

The general framemork I rely on is standard in the literature pioneered by Berry (1994) and Berry et al. (1995) (see also Berry and Haile, 2014). I consider a population of consumers choosing among $J+1$ differentiated products indexed by $j = 0, \dots, J$ in a market t. Product $j = 0$ is referred to as the outside good. Market t is defined by a collection $\left(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \mathbf{p}_t, \boldsymbol{\xi}_t\right)$ of product/market characteristics: $\mathbf{x}_t^{(1)} \equiv (\mathbf{x}_{1t}^{(1)}, \dots, \mathbf{x}_{Jt}^{(1)})$ and $\mathbf{x}_t^{(2)} \equiv (\mathbf{x}_{1t}^{(2)}, \dots, \mathbf{x}_{Jt}^{(2)})$ are two sets of observed characteristics, with $\mathbf{x}_{jt}^{(1)} \in \mathbb{R}^{K_1}$ and $\mathbf{x}_{jt}^{(2)} \in \mathbb{R}^{K_2}$; $\mathbf{p}_t \equiv (p_{1t}, \dots, p_{Jt})$ are observed prices, with $p_{jt} \in \mathbb{R}$; $\boldsymbol{\xi}_t \equiv (\xi_{1t}, \dots, \xi_{Jt})$ are unobservables at the level of product and/or market, with $\xi_{jt} \in \mathbb{R}$, summarizing product differentiation that is not explained by observed characteristics.

Then, I assume a linear index restriction, constraining the way characteristics enter the model. On the one hand, I let $\left(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \mathbf{p}_t, \boldsymbol{\xi}_t\right)$ enter the model through indexes defined as

$$\delta_{jt} \equiv \mathbf{x}_{jt}^{(1)}\boldsymbol{\beta}_1 + \mathbf{x}_{jt}^{(2)}\boldsymbol{\beta}_2 - \alpha p_{jt} + \xi_{jt}, \quad j = 1, \dots, J, \tag{1}$$

where $\alpha > 0$, $\boldsymbol{\beta}_1 \in \mathbb{R}^{K_1}$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^{K_2}$. On the other hand, I let $\mathbf{x}_t^{(2)}$ enter in an unrestricted way, i.e., both through the indexes and directly. This implies that $\mathbf{x}_t^{(1)}$, $\boldsymbol{\xi}_t$ and $\mathbf{p}_t$ will be perfect substitutes and that consumer behavior will be arbitrarily affected by $\mathbf{x}_t^{(2)}$. The logit model does not incorporate characteristics $\mathbf{x}_t^{(2)}$. In the rancom coefficient logit model, characteristics $\mathbf{x}_t^{(2)}$ are those that admit a random coefficient. For the outside good, I set $\delta_{0t} = 0$.

Lastly, I define an empirical model of consumer behavior describing the aggregate behavior of the population of consumers. Different modelings can be used: utility, demand, inverse demand, etc. Here, I specify an inverse market share function $\boldsymbol{\sigma}^{-1} \equiv$

$\left(\sigma_1^{-1}, \ldots, \sigma_J^{-1}\right)$ that, $\mathbf{x}_t^{(2)}$ held fixed, maps from observed market shares $\mathbf{s}_t \equiv (s_{1t}, \ldots, s_{Jt}) \in \Delta_J^+$ to product indexes $\boldsymbol{\delta}_t \equiv (\delta_{1t}, \ldots, \delta_{Jt}) \in \mathbb{R}^J$, where σ_j^{-1} is the inverse market share function for product j and Δ_J^+ denotes the set of non-zero market shares, so that for the outside good, we have $s_{0t} = 1 - \sum_{k=1}^{J} s_{kt}$.[8]

Model. My approach relies on the following inverse market share equations:

$$\sigma_j^{-1}\left(\mathbf{s}_t, \mathbf{x}_t^{(2)}; \{\mu_{ij}\}_{i,j>0}\right) \equiv \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sum_{i \neq j} \mu_{ij} \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right) = \delta_{jt}, \quad j > 0, \tag{2}$$

with

$$\mu_{ij} = \mu\left(\mathbf{d}_{ij,t}^{(2)}; \boldsymbol{\gamma}\right), \qquad \forall i, j > 0,\ i \neq j, \tag{3}$$

$$\mu_{ij} = \mu_{ji}, \qquad \forall i, j > 0,\ i \neq j, \tag{R1}$$

$$\mu_{ij} \geq 0, \qquad \forall i, j > 0,\ i \neq j, \tag{R2}$$

$$\sum_{i \neq j} \mu_{ij} < 1, \qquad \forall j > 0, \tag{R3}$$

where $\mathbf{d}_{ij,t}^{(2)}$ is a vector of measures of proximity between products i and j in characteristics $\mathbf{x}_t^{(2)}$ and $\boldsymbol{\gamma}$ is a parameter vector.

Observe that the inverse market share equations (2) reduce to that implied by the logit model

$$\sigma_j^{-1}(\mathbf{s}_t) \equiv \ln\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt}, \quad j = 1, \ldots, J, \tag{4}$$

when all parameters μ_{ij} shrink to zero. As is well known, the logit model yields counter-intuitive substitution patterns whereby decreases in product j's price reduce the demand for any other product $k \neq j$ by the same percentage, regardless of how close substitutes products are. Thus, Model (2) deviates from the logit model and its restrictive substitution patterns thanks to its positive parameters μ_{ij}. In Section 3, I show how these latter govern substitution patterns between products i and j and allow to obtain flexible substitution patterns.

To obtain richer substitution patterns, the literature has developed demand models that deviate from the logit model in two main ways. The first way introduces unobserved consumer heterogeneity in preferences through random coefficient specifications, as in Berry et al. (1995) (hereafter BLP). The second way uses models based on a nesting structure

[8] That is, $\Delta_J^+ \equiv \left\{\mathbf{s}_t \in (0, \infty)^J : \sum_{j=1}^{J} s_{jt} < 1\right\}$.

(i.e., allocation of products into groups). The most prominent example is the (simple) nested logit model that groups products into exhaustive, mutually exclusive groups (and with a group for the outside good alone) and leads to the following inverse market share equations

$$\sigma_j^{-1}(\mathbf{s}_t;\mu) \equiv \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \mu \ln\left(\frac{s_{jt}}{\sum_{k\in g} s_{kt}}\right) = \delta_{jt}, \quad j = 1,\ldots,J, \tag{5}$$

for a product j in group g. These equations highlight that Model (2) generalizes the inverse market share of the nested logit model by relying on a flexible nesting structure that admits a group for each pair (i, j) of products $i, j > 0$ and a group for the outside good $j = 0$ alone. Hereafter, I refer to parameters μ_{ij} as nesting parameters and Model (2) as the flexible inverse logit (FIL) model as it extends the inverse demand of the logit model flexibly,. Specifically, the FIL model can be viewed as a flexible member of Fosgerau et al. (2021)'s class of models that extends the inverse demand of the nested logit model by allowing any nesting structure. In Section 3, I further show that it is consistent with a representative consumer model and a model of heterogeneous, utility-maximizing consumers.

Equations (3) map the nesting parameters in the characteristic space generated by $\mathbf{x}_t^{(2)}$. They generate substitution patterns depending on how close products are in this space, rather than on their identity, i.e., substitution patterns are in characteristic space, not in product space. Building on Pinkse et al. (2002) and Pinkse and Slade (2004), I express each parameter μ_{ij} as a parametric, pair-invariant function of a vector $\mathbf{d}_{ij,t}^{(2)}$ of measures of proximity between products i and j in characteristics $\mathbf{x}_t^{(2)}$. These mappings allow to explain why consumers are more inclined to substitute among products with similar characteristics, e.g., why consumers whose most preferred car is a BMW tend to switch to another luxury car rather than to a non-luxury car following BMW's price increase. As shown by Gandhi and Houde (2020), BLP also embeds these realistic restrictions on substitution patterns. The key difference is that while in BLP they are determined by random coefficient specifications, here they are determined by the mappings (3).

Restrictions (R1) – (R3) on the nesting parameters μ_{ij} can be motivated in two ways. First, under these restrictions, the inverse market share function $\boldsymbol{\sigma}^{-1}$ in Equations (2) is invertible up to the normalization $\delta_{0t} = 0$ (see Corollary 2 in Appendix C.1). This means that the specified inverse market share function defines a market share function $\boldsymbol{\sigma}$, rather than a correspondence. The normalization is required because otherwise there exists an infinity of vectors $\boldsymbol{\delta}_t$ that can rationalize the vector $\mathbf{s}_t$. It reflects the fact that, before normalization, the market share function is invariant to translation in product indexes $\boldsymbol{\delta}_t$, meaning that differences in $\boldsymbol{\delta}_t$, not their absolute values, determine the market share function. This invertibility result actually extends to a large class of inverse market share functions, including those implied by any additive random utility model without income effect and observed heterogeneity in preferences related to observed individual characteristics and unobserved heterogeneity in preferences through random coefficients (see Proposition 1 in Appendix C.1).

Second, Restrictions (R1) – (R3) have an economic content. As shown in Section 3, they make the matrix of price derivatives of market share symmetric and negative definite. These properties are key for consistency with utility maximization (see Nocke and Schutz, 2017) and for the market share function of each product to be (strictly) decreasing in its own price. Furthermore, they allow for complementarity in demand, defined by a negative cross-price derivative of market share. As a result, the invertibility result above supplements existing results by Berry (1994) and Berry et al. (2013) by not relying on the "connected substitutes" structure and, in particular, by allowing for complementarity in demand.[9]

Before turning to the estimation, two remarks are in order: (i) the random coefficient logit used by BLP also embeds translation invariance of the market share function, and symmetry and negative definiteness of its derivative matrix; (ii) Restrictions (R1) are redundant to the mappings (3), the latter implying the former since $\mathbf{d}^{(2)}_{ij,t} \equiv \mathbf{d}^{(2)}_{ji,t}$.

Identification and Estimation. The market-level data I rely on are the same as BLP. Consider having data $\{s_{jt}, p_{jt}, \mathbf{x}^{(1)}_{jt}, \mathbf{x}^{(2)}_{jt}\}$ on market shares, prices and product characteristics for $j = 1, \ldots, J$ products in $t = 1, \ldots, T$ markets.

Combining Equations (2) and (3) allows to obtain each unobserved characteristic term ξ_{jt} as a parametric function of data and demand parameters $(\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma})$

$$\xi_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sum_{i\neq j} \mu\left(\mathbf{d}^{(2)}_{ij,t}; \boldsymbol{\gamma}\right) \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right) - \mathbf{x}^{(1)}_{jt}\boldsymbol{\beta}_1 - \mathbf{x}^{(2)}_{jt}\boldsymbol{\beta}_2 + \alpha p_{jt}. \quad (6)$$

Following the literature, the ξ_{jt} terms are structural error terms, as they summarize all the product/market characteristics observed by consumers and firms but not by the modeler. Furthermore, product characteristics $\mathbf{x}^{(1)}_{jt}$ and $\mathbf{x}^{(2)}_{jt}$ are assumed to be exogenous (i.e.,

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

uncorrelated with ξ_{jt}), whereas prices and market shares are considered as endogenous.[10]

Then, provided that there exist instruments $\mathbf{z}_t$ for prices and market shares, we can estimate demand parameters based on the following conditional moment restrictions (Berry, 1994)

$$\mathbb{E}[\xi_{jt}|\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \mathbf{z}_t] = 0, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T \tag{7}$$

where $\boldsymbol{\gamma}$ is such that Restrictions (R1) – (R3) are satisfied.

In Section 4, I provide details about estimation, identification, and implementation. However, I can make three comments at this stage. First, I will only require parametric instrumental variable regressions for estimation. This is because my approach relies on an inverse market share function, which has an explicit formula known up to some parameters to be estimated. This contrasts with BLP that uses the random coefficient logit model for which there is no explicit inversion formula such as in Equation (6), and in turn, prevents us from using regressions.

Second, the structural interpretation of the error terms help build intuition for identification using instruments. As in the literature (see Berry and Haile, 2014), any variable that induces exogenous variation in choice sets will be good candidates as instruments. As a result, I will use standard instruments of the literature, thereby relying only on conventional sources of empirical variation to identify demand parameters.

Third, I will impose Restrictions (R1) – (R3) during estimation. This builds on the literature on the econometrics of shape restrictions (Matzkin, 1994; Chetverikov et al., 2018) that shows how imposing restrictions during estimation help improve the estimation. At the same time, this makes inference harder. A future version of this paper will propose a method to handle this issue.

3 Theory: Substitution Patterns and Utility

In this section, I provide theoretical details about my approach. I first describe how it can accommodate rich substitution patterns in characteristic space. Then, I show that it is consistent with utility maximization.

[10] For example, consider a price competition model. Prices are endogenous because firms take into account both observed and unobserved characteristics when they set their prices. Furthermore, market shares are endogenous as they are determined by a full system of equations involving the entire vectors of endogenous prices and unobserved characteristics, and consumers choose products while potentially considering the unobserved characteristics.

3.1 Substitution Patterns

Substitution patterns refer to how consumers substitute among products following a price increase. They can be captured by the matrix of own- and cross-price derivatives of market share: own-price derivatives express how much a one-unit product's price increase would cause a drop in its own market share, whereas cross-price derivatives indicate where the lost demand goes. Absent income effect, the matrix of price derivatives of market share is equivalent to the Slustky matrix.

Formally, the Slutsky matrix is given by

$$\left[\frac{\partial \sigma_i(\boldsymbol{\delta}_t)}{\partial p_{jt}}\right] = -\alpha \left[\frac{\partial \sigma_i(\boldsymbol{\delta}_t)}{\partial \delta_{jt}}\right], \tag{8}$$

where σ_i denotes the market share function for product i. Let $\boldsymbol{\sigma}^{-1}(\mathbf{s}_t) = \boldsymbol{\delta}_t$. Then,[11]

$$\left[\frac{\partial \sigma_i(\boldsymbol{\delta}_t)}{\partial p_{jt}}\right] = -\alpha \left[\frac{\partial \sigma_i^{-1}(\mathbf{s}_t)}{\partial s_{jt}}\right]^{-1}. \tag{9}$$

For the FIL model (2), we have

$$\frac{\partial \sigma_i^{-1}(\mathbf{s}_t)}{\partial s_{jt}} = \begin{cases} \dfrac{1}{s_{0t}} + \dfrac{1 - \sum_{i \neq j} \mu_{ij}}{s_{jt}} + \sum_{i \neq j} \dfrac{\mu_{ij}}{s_{it} + s_{jt}}, & \text{if } i = j, \\[2ex] \dfrac{1}{s_{0t}} + \dfrac{\mu_{ij}}{s_{it} + s_{jt}}, & \text{if } i \neq j, \end{cases} \tag{10}$$

where μ_{ij} is mapped in characteristic space according to Equations (3).

Restrictions (R1) – (R3) have implications for substitution patterns. Restrictions (R1) imply Slutsky symmetry, while Restrictions (R2) and (R3) imply Slutsky negative definiteness.[12] Absent income effect, these properties are key for the FIL model to be consistent with the maximization of a quasi-linear utility function (see, e.g., Nocke and Schutz,

[11] Let $\boldsymbol{\sigma}^{-1}(\mathbf{s}_t) = \boldsymbol{\delta}_t$. Then, use that $\boldsymbol{\sigma}^{-1}$ is invertible, with inverse $\boldsymbol{\sigma}$, and apply the chain rule (Simon and Blume, 1994, Theorem 14.4).

[12] Restriction (R1) implies that the matrix (10) is symmetric, which implies that its inverse, the Slustky matrix (8), is symmetric as well. The matrix (10) is the sum of two matrices, where the first one is positive semi-definite and the second one, by Restrictions (R2) and (R3), is positive definite as it is a symmetric, strictly diagonally dominant matrix with positive diagonal entries (Horn and Johnson, 2012, Theorem 6.1.10.) Then the matrix (10) is positive definite and so is its inverse. As $\alpha > 0$, this implies that the Slustky matrix (8) is negative definite.

2017).[13] Furthermore, they entail that the demand of each product is strictly decreasing in its own price (law of demand). Lastly, they do not restrict products to be substitutes in demand.

The mappings (3) also have implications for substitution patterns. To see this, let $y_{jt} \equiv (s_{jt}, \mathbf{x}^{(2)}_{jt})$ and $\mathbf{y}_{-jt} \equiv (y_{1t}, \ldots, y_{j-1,t}, y_{j+1,t}, \ldots, y_{Jt})$. Then, the FIL model (2) mapped in characteristic space can be rewritten as follows (see Appendix C.3 for details)

$$\sigma_j^{-1}\left(y_j, \mathbf{y}_{-j}; \boldsymbol{\gamma}\right) = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sum_{i=1}^{J} \mu\left(\mathbf{d}^{(2)}_{ij,t}; \boldsymbol{\gamma}\right) \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right) + C_t, \quad j = 1, \ldots, J, \tag{11}$$

where $C_t \in \mathbb{R}$ is a market-specific constant and where the sum is over all products including product j itself.

The inverse market share function (11) for product j exhibits three key features. First, it is product-invariant and only depends on market shares $\mathbf{s}_t$ and characteristics $\mathbf{x}^{(2)}_t$

$$\sigma_j^{-1}\left(y_j, \mathbf{y}_{-j}; \boldsymbol{\gamma}\right) = \sigma_k^{-1}\left(y_j, \mathbf{y}_{-j}; \boldsymbol{\gamma}\right) = \sigma^{-1}\left(y_j, \mathbf{y}_{-j}; \boldsymbol{\gamma}\right), \quad \text{for any } j \neq k. \tag{12}$$

Second, it does not depend on the ordering of competing products $k \neq j$ but only on their market shares $\mathbf{s}_{-jt}$ and characteristics $\mathbf{x}^{(2)}_{-jt}$:

$$\sigma^{-1}\left(y_j, \mathbf{y}_{-j}; \boldsymbol{\gamma}\right) = \sigma^{-1}\left(y_j, \mathbf{y}_{\rho(-j)}; \boldsymbol{\gamma}\right) \tag{13}$$

where $\rho(-j)$ is any permutation of $-j \equiv (1, \ldots, j-1, j+1, \ldots, J)$. These two features imply that it is not the identity of products but their characteristics and popularity that determine substitution patterns.

Third, the inverse market share function (11) does not depend on the absolute value of the characteristics but on their relative values. Formally, it is invariant to translation in $\mathbf{x}^{(2)}_t$:

$$\sigma^{-1}\left(y_j + (0, c), \mathbf{y}_{-j} + (\mathbf{0}_{J-1}, c\mathbf{1}_{J-1}); \boldsymbol{\gamma}\right) = \sigma^{-1}\left(y_j, \mathbf{y}_{-j}; \boldsymbol{\gamma}\right), \quad \text{for all } c \in \mathbb{R}, \tag{14}$$

where $\mathbf{0}_{J-1} \in \mathbb{R}^{J-1}$ is a vector of zeros and $\mathbf{1}_{J-1} \in \mathbb{R}^{J-1}$ is a vector of ones. It means that it is not the level of product characteristics but their closeness in characteristic space that determine substitution patterns.

These three features reveal how the mappings (3) generate substitution patterns deter-

[13] Note that I can use Nocke and Schutz (2017)'s results because the market share function is continuously differentiable.

mined by the proximity of products in the characteristic space generated by $\mathbf{x}_t^{(2)}$ – rather than their identity, i.e., substitution patterns are in product space – rather than in characteristic space. [14]

Equation (11) help build intuition on how my approach generates rich substitution patterns. Observe first that the nesting structure is just a valuable way to fully parametrize the matrix of (inverse) market share derivatives. It means, by contrast to the nested logit model, that the nesting structure of the FIL model does not require the modeller to take a stand on the relevant dimensions along which groups can be defined and its implied substitution patterns are not constrained by a priori restriction such as product segmentation.

Then, note that the FIL model is flexible in the sense of Diewert (1974) in a large class of well-defined inverse market share functions, encompassing all those implied by any additive random utility model without income effect and observed heterogeneity in preferences related to observed individual characteristics and unobserved heterogeneity in preferences through random coefficients. That is, the FIL model can match the vector of market shares as well as any matrix of own- and cross-price elasticities implied by an inverse market share function of this class.[15]

This result is formally stated in Proposition 3 in Appendix C.2. Its proof can be sketched in two steps. The first step uses the invertibility of the FIL model to show that there always exists a vector $\boldsymbol{\delta}_t$ of indexes that equates the vector $\mathbf{s}_t$ of observed shares to the vector $\boldsymbol{\sigma}$ of predicted shares. The second step shows that the FIL model can match any own- and cross-price elasticities, which, once the market shares are matched, is equivalent to match the own- and cross-price derivatives. Intuitively, we can match the cross-price derivative $\partial\sigma_i(\boldsymbol{\delta}_t)/\partial p_{jt}$ by appropriately choosing the value of the nesting parameter μ_{ij}. Once this is done, all own-price elasticity $\partial\sigma_j(\boldsymbol{\delta}_t)/\partial p_{jt}$ are automatically matched because, as for any unit-demand model

$$\frac{\partial\sigma_j(\boldsymbol{\delta}_t)}{\partial p_{jt}} = -\sum_{k\neq j}\frac{\partial\sigma_k(\boldsymbol{\delta}_t)}{\partial p_{jt}}. \tag{15}$$

[14]The first two features are referred to as symmetry and anonymity. Therefore, my approach closely relates to Compiani (2021) who uses anonymity to reduce the dimensionality of his non-parametric estimation, and to Gandhi and Houde (2020) who use anonymity and symmetry of the linear-in-characteristics random coefficient logit model to construct new approximations of the optimal instruments.

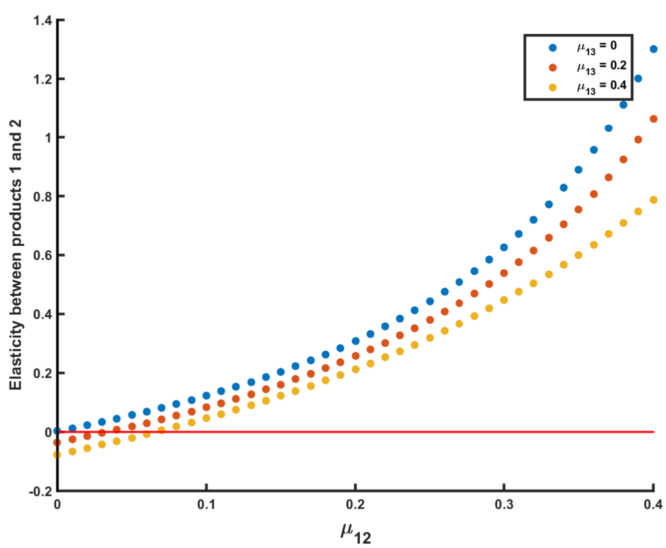
[15]A demand system is flexible in the sense of Diewert (1974) if it is able to provide a first-order approximation to any theoretically grounded demand system at a point in price space. Equivalently, flexibility can also be viewed as the ability of the (direct or indirect) utility function to provide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

Note that flexibility in the sense of Diewert (1974) means that we can match price elasticities in a single market. That is, given a single market, we can conveniently choose values for the parameters μ_{ij} to match the price elasticities implied by a nested logit model. However, flexibility does not hold when we consider multiple markets, i.e., a flexible model cannot match the price elasticities in all markets simultaneously. By contrast, simulations in Section 5 show how the mappings (3) can yield rich substitution patterns in all market simultaneously.

This discussion has highlighted the role of the nesting parameters in driving substitution patterns. To obtain further intuition into this, I consider the following stylized example.

Example 1. Let $J = 3$ and assume that $\mathbf{x}_t^{(1)} = (0.3, 0.25, 0.2)$, $\mathbf{p}_t = (1.1, 1, 0.9)$, $\xi_{jt} \sim \mathcal{N}(0, 0.15^2)$, $\delta_{jt} = -5 - p_{jt} + x_{jt}^{(1)} + \xi_{jt}$. Set $\mu_{23} = 0.4$ and let μ_{12} and μ_{13} vary. The scatter plot below shows that higher values for μ_{12} implies higher elasticity between products 1 and 2. By varying the value of μ_{13}, it also shows that the relationship between products 1 and 2 (substitutes/complements) is affected by product 3.

Figure 1: The nesting parameter μ_{ij} governs substitution patterns between products i and j



The scatter plot shows the cross-price demand elasticity of product 1 with respect to product 2 as a function of μ_{12} for different values of μ_{13}. The red horizontal line correspond to the threshold between complementarity and substitutability in demand.

Example 1 shows that the nesting parameter μ_{ij} governs substitution patterns between products i and j. First, the higher the value for μ_{ij}, the higher the cross-price elasticity η_{ij}.

Then, the FIL model allows for complementarity in demand. Lastly, whether products i and j are complements or substitutes does not uniquely depend on μ_{ij} but also on μ_{ik}, $k \neq j$, $k \neq i$. This finding is consistent with the theoretical result whereby whether two products are complements or substitutes in demand depends on the relation of the two products to the other products (Samuelson, 1974).

3.2 Utility Maximization

Recall that my approach relies on the FIL model defined in Equations (2). As shown above, it is an inverse market share function that is consistent with utility maximization. I here show that it can be derived from a representative consumer model and a model of heterogeneous, utility-maximizing consumers.

Representative Consumer. Building on Fosgerau et al. (2021), I first show that the FIL model is consistent with a representative consumer whose utility function describes the aggregate behavior of a population of possibly heterogeneous consumers.

Consider a representative consumer choosing among $J + 1$ differentiated products and a homogeneous numéraire good in a market t. Let v_{jt} be the quality of the differentiated product $j = 0, \ldots, J$ in market t and p_{jt} be its price. In the empirical framework, we have $v_{jt} = \mathbf{x}_{jt}^{(1)}\boldsymbol{\beta}_1 + \mathbf{x}_{jt}^{(2)}\boldsymbol{\beta}_2 + \xi_{jt}$. The price of the numéraire good is normalized to 1 and the representative consumer's income y is assumed to be sufficiently high for the consumption of the numéraire good be positive.

Let $\mu_j \equiv 1 - \sum_{i \neq j} \mu_{ij}$ for all $j = 1, \ldots, J$. The FIL model is consistent with a representative consumer who chooses market shares $s_{jt} \geq 0$, $j = 0, \ldots, J$ of the differentiated products and a quantity $z_t \geq 0$ of the numéraire good to maximize her direct utility function u defined by

$$u(z_t, \mathbf{s}_t) = \alpha z_t + \sum_{j=0}^{J} v_{jt} s_{jt} - \left[s_{0t} \ln(s_{0t}) + \sum_{j=1}^{J} s_{jt} \left(\mu_j \ln(s_{jt}) + \sum_{i \neq j} \mu_{ij} \ln(s_{it} + s_{jt}) \right) \right] \tag{16}$$

if $\sum_{j=0}^{J} s_{jt} = 1$ and $u = -\infty$ otherwise, subject to her budget constraint

$$\sum_{j=0}^{J} p_{jt} s_{jt} + z_t \leq y. \tag{17}$$

As for the logit and nested logit models (Anderson et al., 1988; Verboven, 1996b), utility (16) embeds two effects. The first effect, given by $\alpha z_t + \sum_{j=0}^{J} v_{jt} s_{jt}$, captures the utility that the representative consumer derives from consuming the differentiated products and the numéraire in the absence of interaction (in utility) among them. The second effect, given by the expression into brackets, models its preference for variety. Indeed, if utility (16) were only given by $\alpha z_t + \sum_{j=0}^{J} v_{jt} s_{jt}$, then the representative consumer would maximize her utility by choosing the product j with the highest utility v_{jt}. Overall, the representative consumer chooses a positive quantity of every product while trading-off variety against quantity. With this representation, similarly to the nested logit model, each parameter μ_{ij} measures taste for variety over products i and j; the intuition being that $\mu_{ij} > 0$ makes the representative consumer more willing to choose products i and j. Furthermore, the parameter $\alpha > 0$ expresses consumer' price sensitivity or its marginal utility of income, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ capture consumer' taste for characteristics $\mathbf{x}_{jt}^{(1)}$ and $\mathbf{x}_{jt}^{(2)}$, respectively.

A Model of Consumer Heterogeneity. The FIL model can also be derived as specific instance of the large class of utility models studied by Allen and Rehbeck (2019). To see this, note first that Restrictions (R1) – (R3) imply that the term into brackets of utility (16) is a strictly concave function of $\mathbf{s}$ that do not depend on $\boldsymbol{\delta}$. Then, by Allen and Rehbeck (2019), utility (16) can be derived, after an aggregation across consumers, from a model of heterogeneous, utility-maximizing consumers. Importantly, we do not need to know and identify the distribution of heterogeneity to estimate demand parameters.

As a result, the FIL model allows for (unobserved) consumer heterogeneity in preferences modelled by the second term of utility (16). However, as it stands, it does not allow for observed heterogeneity in preferences related to observed individual characteristics or unobserved heterogeneity in preferences through random coefficients. With this representation, the parameters μ_{ij} control for the distribution of preferences in the population of consumers.

A Remark on the μ_{ij}'s. The nesting parameters μ_{ij} are structural parameters, as they describe consumers' preferences and they are invariant to changes in regulation (e.g., taxes) and in firms' strategy (e.g., pricing strategies, product characteristics) (see Hurwicz, 1966). However, a question is whether or not the nesting parameters mapped in characteristic space (3) are structural. It is clear that they are not structural, since they are no longer invariant to changes in product characteristics by firms. However, parameters in $\boldsymbol{\gamma}$, which parametrize the mappings, are structural and can be interpreted as controlling for the distribution of valuation for product characteristics $\mathbf{x}_t^{(2)}$ in the population, as the random coefficients in BLP.

4 Empirical Strategy

4.1 Estimation and Identification

Estimation. Recall that estimation using market-level data is based on the conditional moment restrictions (7). These conditional moment restrictions lead to the following unconditional moment restrictions

$$\mathbb{E}[\xi_{jt}\tilde{\mathbf{z}}_t] = 0, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T \tag{18}$$

where $\tilde{\mathbf{z}}_t = \{\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, h_{jt}(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \mathbf{z}_t)\}$, with h_{jt} functions of $\mathbf{x}_t^{(1)}$, $\mathbf{x}_t^{(2)}$ and $\mathbf{z}_t$, and where I recall that ξ_{jt} is given by Equation (6) and $\boldsymbol{\gamma}$ satisfies Restrictions (R1) – (R3).

Identification. Identification of the FIL model amounts to identifying demand parameters $\{\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma})$. As explained above, the FIL model boils down to an IV regression where prices and log-share terms are endogenous. Therefore, the main identification assumption is the existence of instruments $\mathbf{z}_t$, that is, variables that induce enough independent exogenous variation in each of these endogenous variables. The discussion about identification thus reduces to the question about what sources of empirical variation help learn about demand parameters.

Consider first the vector of utility indexes $\boldsymbol{\delta}_t$. It is easy to identify, as for two products with identical characteristics, higher market shares imply higher utility indexes. Formally, as shown in Proposition 1, assuming $\delta_{0t} = 0$ and given a value for μ_{ij}'s, there is a one-to-one mapping between the vector of utility indexes and the vector of market shares.

Regarding the parameters α, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ entering the utility indexes, their identification is also fairly simple. As is well known in the literature (see, e.g., Berry, 1994; Berry et al., 1995; Berry and Haile, 2014, etc.), it requires dealing with price endogeneity, which is done by using valid supply-side instruments, i.e., cost shifters and/or markup shifters. The first set of instruments includes Hausman instruments, i.e., prices in other markets (Hausman et al., 1994; Nevo, 2001). The second set involves BLP instruments, i.e., functions of the characteristics of competing products (Berry et al., 1995; Gandhi and Houde, 2020), as well as market shocks such as mergers (Miller and Weinberg, 2017).

Turn now to the nesting parameters μ_{ij}, which, as mentioned above, govern substitution patterns between products i and j. Their identification is more tricky since it requires exogenous variation in the relative popularity of product j with respect to product i. It means that we need instruments that reveal the substitution patterns among products. Variables that generate exogenous variation in the choice set (i.e., changes in prices, product characteristics and number of products) are therefore good candidates as instruments.

To gain further insights into the identification of the μ_{ij}'s, I consider a stylized example with $J = 3$ products and a single market t, where the model of consumer behavior is a FIL model and the model of firm conduct is price competition with three single-product firms. I assume that the modeler observes data on prices and market shares. For the example, I set $s_{1t} = 0.075$, $s_{2t} = 0.0125$, $s_{3t} = 0.01$ and $p_{1t} = p_{2t} = p_{3t} = 1$. I want to understand how different sources of empirical variation help identification of μ_{12}, μ_{13} and μ_{23}. For the sake of clarity, I set $\mu_{13} = \mu_{23} = 0.3$ and study identification of μ_{12}. Specifically, I consider two sets of empirical variation: (i) a 10% increase in product 1's cost, $\Delta c_1 = 10\%$; (ii) a merger between firms 1 and 2.

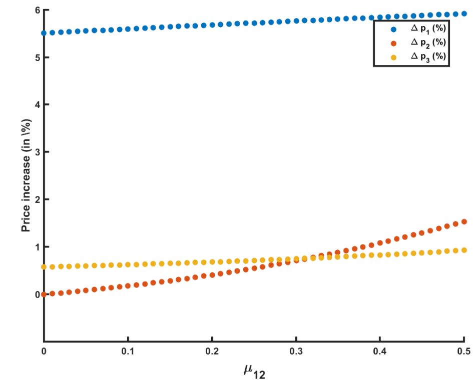
Figure 2 show how prices (top panels) and relative shares (bottom panels) are affected by the cost increase (left panels) and the merger (right panels), depending on the level of μ_{12}. It illustrates monotonic relationships between the variation in prices and μ_{12} on one hand, and between the variation in relative shares and μ_{12} on other hand. For example, a 10% increase in product 1's cost leads to 5.5% increase in its own price when $\mu_{12} = 0$ and a 6% increase when $\mu_{12} = 0.5$. Thereby, the way prices and relative shares change with product 1's cost increase or with the merger drives the estimates of μ_{12}.

Lastly, consider the mappings. Identification requires finding a unique vector of parameters $\boldsymbol{\gamma}$ such that $\mu_{ij} = \mu\left(\mathbf{d}_{ij,t}^{(2)}\right)$ for all $i \neq j$. For example, in the one-dimensional mapping described where the mapping is specified using a Bernstein polynomial, the parameter vector $\boldsymbol{\gamma}$ can be obtained as the OLS estimates of a regression of μ_{ij} on polynomial basis. The identification assumptions are therefore the same as in a OLS setting.

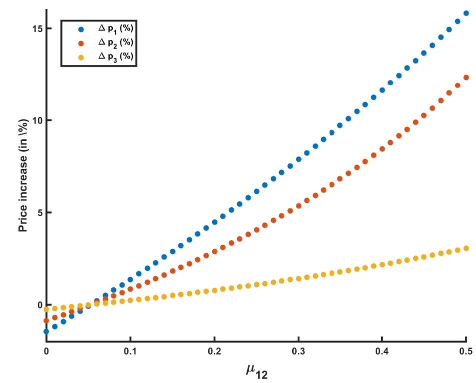
4.2 Implementation

I conclude this section by providing a user guide to estimating substitution patterns using my approach. For simplicity, I present the simplest case, also used in the simulations of the

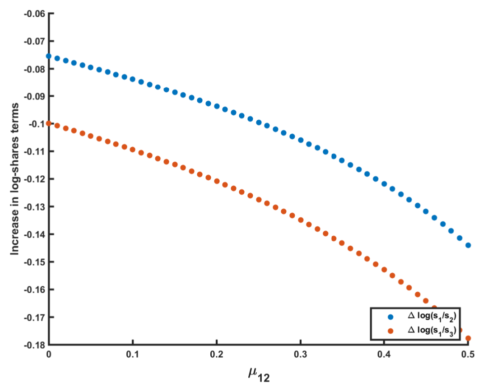
Figure 2: Insights on identification



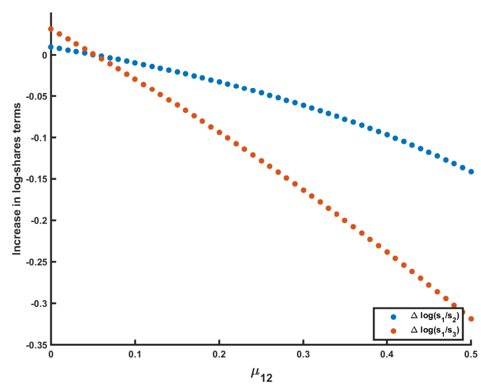
(a) Cost shifter and Prices



(c) Merger and Prices



(b) Cost shifter and Relative Shares



(d) Merger and Relative Shares

next section, where the mappings are uni-dimensional, i.e., when $x_{jt}^{(2)} \in \mathbb{R}$.

Step 0. Preliminary choices. Consider mapping the μ_{ij}'s in the characteristic space generated by $x_{jt}^{(2)} \in [0, 1]$. The modeler has two choices to make.

1. *Measure of proximity.* There exist different possible measures of proximity (based on the Euclidian distance, absolue value, etc.). I choose

$$d_{ij,t}^{(2)} = 1 - |x_{it} - x_{jt}|, \tag{19}$$

   where $d_{ij,t}^{(2)} = 0$ when there is minimal proximity and $d_{ij,t}^{(2)} = 1$ when there is maximal proximity.

2. *Specification of μ.* There are different ways of specifying μ as a function of $d_{ij,t}^{(2)}$. As

mentioned above, I use a Bernstein polynomial of order D in $d_{ij,t}^{(2)}$

$$\mu_{ij} = \mu\left(d_{ij,t}^{(2)}; \boldsymbol{\gamma}\right) = \sum_{k=0}^{D} \gamma_k \binom{D}{k} \left(d_{ij,t}^{(2)}\right)^k \left(1 - d_{ij,t}^{(2)}\right)^{D-k}, \tag{20}$$

so that $\boldsymbol{\gamma} \equiv (\gamma_0, \ldots, \gamma_M)$. In practice, we have to choose the order D of the polynomial (e.g., by using cross-validation).

Step 1. Computation of instruments $\mathbf{z}_t$. Building on insights from Newey (1990), Belloni et al. (2012) and Gandhi and Houde (2020), we can construct instruments as the predicted values from lasso regressions of each endogenous variable on functions of the exogenous variables of the model.

For example, for the price variable, we can consider two sets of exogenous variables:

1. Product characteristics $\mathbf{x}_{jt}^{(1)}$ and $\mathbf{x}_{jt}^{(2)}$, cost shifters $\mathbf{z}_{jt}$, and possibly interactions and polynomials.

2. Differentiation instruments (Gandhi and Houde, 2020) in $\mathbf{x}_{jt}^{(1)}$, $\mathbf{x}_{jt}^{(2)}$ and $\mathbf{z}_{jt}$:

$$\sum_i (u_{it} - u_{jt})^k, \quad \text{and} \quad \left(\sum_i (u_{it} - u_{jt})\right)^k, \tag{21}$$

where $u_{jt} \in \{\mathbf{x}_{jt}^{(1)}, \mathbf{x}_{jt}^{(2)}, \mathbf{z}_{jt}\}$ and the sums are (i) over all products (except j), (ii) over products of the same firm and (iii) over products of rival firms.

The first set is intended to proxy for the marginal cost part of the price, whereas the second set should proxy for the markup part of the price.

Step 2. Constraints. The modeler must define the restrictions she wants to impose on the nesting parameters μ_{ij}. As mentioned above, Restrictions (R1) are already imposed via the mappings.

As shown by Chak et al. (2005), the non-negativity restrictions (R2) on μ_{ij} can be converted into a set of non-negativity restrictions on the coefficients γ_k of the Berstein polynomial:

$$\gamma_k \geq 0, \quad \text{for all} \quad k = 0, \ldots, D. \tag{22}$$

For Restrictions (R3), there is no such a result. Then, I directly enforce them. Lastly, the modeler may also want to restrict products to be substitutes in demand. For example, we may believe that cars are substitutes. I have empirically found that imposing constraints (R2) and $\sum_{i \neq j} \mu_{ij} \leq 1/2$ help enforcing substitutability.[16]

Step 3. Constrained generalized method of moments (GMM). I employ the iterated GMM estimator, while imposing the restrictions defined just above. Let $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma})$. Let $N = J \times T$, $\boldsymbol{\xi} = (\xi_{11}, \ldots, \xi_{jt}, \ldots, \xi_{JT})'$ and $\mathbf{z} = (\mathbf{z}_{11}, \ldots, \mathbf{z}_{jt}, \ldots, \mathbf{z}_{JT})'$.

- *Stage 1.* Obtain $\widehat{\boldsymbol{\theta}}^1$ and $\widehat{\boldsymbol{\xi}}^1$ where

$$\widehat{\boldsymbol{\theta}}^1 = \arg\min_{\boldsymbol{\theta}} \left(\frac{\boldsymbol{\xi}'\mathbf{z}}{N}\right) \left(\frac{\mathbf{z}'\mathbf{z}}{N}\right)^{-1} \left(\frac{\mathbf{z}'\boldsymbol{\xi}}{N}\right) \quad \text{subject to restrictions.} \tag{23}$$

  and $\widehat{\boldsymbol{\xi}}^1$ are the residuals.

- *Stage* $s > 1$. Let $\mathbf{D} \equiv \operatorname{diag}(\widehat{\boldsymbol{\xi}}^{s-1})$. Obtain $\widehat{\boldsymbol{\theta}}^s$ and $\widehat{\boldsymbol{\xi}}^s$, where

$$\widehat{\boldsymbol{\theta}}^s = \arg\min_{\boldsymbol{\theta}} \left(\frac{\widehat{\boldsymbol{\xi}}^{s-1\prime}\mathbf{z}}{N}\right) \left(\frac{\mathbf{z}'\mathbf{D}\mathbf{z}}{N}\right)^{-1} \left(\frac{\mathbf{z}'\widehat{\boldsymbol{\xi}}^{s-1}}{N}\right) \quad \text{subject to restrictions.} \tag{24}$$

  and $\widehat{\boldsymbol{\xi}}^s$ are the residuals.

- Repeat Stage s until convergence defined as $|\widehat{\boldsymbol{\theta}}^s - \widehat{\boldsymbol{\theta}}^{s-1}| < 1e - 05$. Obtain demand estimates $\widehat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\gamma}})$.

Step 4. Estimated Price-Elasticities and Markups. For a market t, the estimated matrix of price elasticities of market share function is computed as follows.

$$-\hat{\alpha} \left[\operatorname{diag}(\hat{\mathbf{s}}_t)\right]^{-1} \left[\frac{\partial \sigma_i^{-1}(\hat{\mathbf{s}}_t)}{\partial s_{jt}}\right]^{-1} \left[\operatorname{diag}(\mathbf{p}_t)\right], \tag{25}$$

where $\hat{\mathbf{s}}_t$ denotes the vector of predicted market shares and $\partial \sigma_i^{-1}(\hat{\mathbf{s}}_t) / \partial s_{jt}$ is defined by Equation (10) with $\mu_{ij} = \mu\left(\mathbf{x}_t^{(2)}; \hat{\boldsymbol{\gamma}}\right)$.

When combined with a model of firm conduct, demand estimates can be used to back out estimated marginal costs and thus markups (Bresnahan, 1989; Berry et al., 1995; Nevo,

[16] Even when these constraints are enforced, there still exist cases where complementarity can occur. However, it happens very seldom.

2001). Following the literature, I consider a static oligopolistic price competition model with F multiproduct firms. That is, in each market t, each firm $f = 1, \ldots, F$ produces a set of products $\mathcal{J}_f$ and chooses the prices p_{jt} for its products $j \in \mathcal{J}_f$ to maximize its profit function given by

$$\Pi_{ft} = \sum_{j \in \mathcal{J}_f} (p_{jt} - c_{jt})\, \sigma_j(\mathbf{p}_t), \tag{26}$$

where σ_j is the demand function for product j implied by the FIL model and where c_{jt} is its marginal cost. Assuming that a pure-strategy Nash equilibrium in prices exists, we can use the associated first-order conditions to back out the estimated marginal cost $\hat{c}_{jt}$.

Then, the estimated relative markup (in percentage) of product j in market t is computed as follows

$$100 \times \frac{p_{jt} - \hat{c}_{jt}}{p_{jt}}. \tag{27}$$

Step 5. Merger Simulation. Assuming that a pure-strategy Nash equilibrium in prices exists, demand and marginal costs estimates can be used together with the associated post-merger first-order conditions to compute the post-merger equilibrium prices (Baker and Bresnahan, 1985; Nevo, 2000a).

5 Monte-Carlo Simulations

This section provides Monte-Carlo simulations. They have two purposes. First, they show that my approach works well with samples of moderate sizes – in simulations I use the same amount of data as the pseudo-real data set from Nevo (2000b).

Second, they help compare my approach to BLP in terms of price elasticities, markups, and merger's price effect. To this end, I take BLP that uses a normal random coefficient on exogenous characteristics, like in the seminal paper by BLP, as the benchmark.

5.1 Data Generating Processes

Both BLP and my approach combine a model of consumer behavior and a model of firm conduct. The model of consumer behavior is a static random coefficient logit (RCL) model in BLP,[17] and the FIL model mapped in characteristic space in my approach. In both

[17] See Berry et al. (1995); Nevo (2000b); Conlon and Gortmaker (2020) for details.

approaches, the model of firm conduct is a static oligopolistic price competition model with multiproduct firms.

FIL Model. In the simulations, I consider FIL models mapped in characteristic space defined by Equations (2) – (3) with linear, one-dimensional mappings in the characteristic space generated by $x_{jt}^{(2)} \in \mathbb{R}$[18]

$$\mu_{ij} = \gamma_0 + \gamma_1 \left(1 - |x_{it}^{(2)} - x_{jt}^{(2)}|\right), \tag{28}$$

and linear indexes given by

$$\delta_{jt} = \beta_0 + \beta x_{jt}^{(2)} - \alpha p_{jt} + \xi_{jt}. \tag{29}$$

RCL Model. In the simulations, I consider RCL models with a single random coefficient on an exogenous continuous characteristics $x_{jt}^{(2)} \in \mathbb{R}$. The conditional indirect utility of a consumer n in market t from choosing product $j = 1, \dots, J$ is given by

$$u_{njt} = \beta_0 + \beta_n x_{jt}^{(2)} - \alpha p_{jt} + \xi_{jt} + \varepsilon_{njt}, \tag{30}$$

where the ε_{njt} are distributed i.i.d. type I extreme value and where the utility from choosing the outside good $j = 0$ is $u_{n0t} = \varepsilon_{n0t}$, for all markets $t = 1, \dots, T$. Each consumer n chooses one unit of the product that provides her the highest utility. Then, the market share of product j in market t is computed as the probability that product j provides the highest utility across all products in market t.

Simulations and estimations of the RCL model make use of the Python package `PyBLP` by Conlon and Gortmaker (2020), which implements the best practices for estimating RCL models using BLP. In particular, I construct instruments into two steps. First, I use the differentiation instruments developed by Gandhi and Houde (2020). Second, I update to the optimal instruments of Chamberlain (1987).

Price Competition Model. I consider a price competition model as defined in the previous section. Assuming that a pure-strategy Nash equilibrium in prices exists, prices $\mathbf{p}_t$ and market shares $\mathbf{s}_t$ are determined by the associated first-order conditions.

[18]Note that estimation of the FIL model when the data generating process is from BLP uses Bernstein polynomials of order $D > 1$.

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

$$c_{jt} = \psi_0 + \psi_x x_{jt}^{(2)} + \psi_w w_{jt} + \omega_{jt},$$

where w_{jt} is a cost shifter affecting only marginal costs, and ω_{jt} is an unobserved cost component.

Data Generating Processes (DGPs). In each DGP, I construct 50 Monte Carlo datasets. Each dataset describes $T = 100$ markets with $J = 25$ products and $F = 5$ firms, each one producing 5 products. Each market t is characterized by $(s_{jt}, p_{jt}, x_{jt}, \xi_{jt}, w_{jt}, \omega_{jt})_{j=1,\dots,J}$, where x_{jt} and w_{jt} are drawn from two independent standard uniform distributions, and where the vector of error terms

$$\begin{bmatrix} \xi_{jt} \\ \omega_{jt} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.15^2 & 0.05 \\ 0.05 & 0.15^2 \end{bmatrix} \right). \tag{31}$$

In all DGPs, I set $\psi_0 = 2$ and $\psi_x = \psi_w = 1$.

DGP 1.BLP with a Normal Random Coefficient. I consider two specifications:

(a) Moderate unobserved heterogeneity: $\alpha = 1$, $\beta_0 = 3$ and $\beta_n \sim \mathcal{N}(3, 6^2)$.

(b) Substantial unobserved heterogeneity: $\alpha = 1$, $\beta_0 = 3$ and $\beta_n \sim \mathcal{N}(3, 12^2)$.

DGP 2. BLP with a Log-normal Random Coefficient. I consider two specifications:

(a) $\beta_0 = 3$ and $\beta_n \sim 3 + \log \mathcal{N}(0, 3)$.

(b) $\beta_0 = 6$ and $\beta_n \sim 3 + \log \mathcal{N}(0, 3.5)$.

DGP 3. Gentzkow Model with interaction parameter Γ. [19] I consider one specification.

(a) $\alpha = 1$, $\beta_0 = 3$, $\beta_n \sim \mathcal{N}(3, 3^2)$, and $\Gamma = 6$.

DGP 4. FIL Model with a Linear μ. I consider two specifications:

(a) $\alpha = 1/4$, $\beta_0 = \beta = 1/4$ and $\mu_{ij} = 0.04 \left(1 - |x_{it}^{(2)} - x_{jt}^{(2)}|\right)$.

(b) $\alpha = 1/2$, $\beta_0 = \beta = 1/2$ and $\mu_{ij} = 0.02 + 0.02 \left(1 - |x_{it}^{(2)} - x_{jt}^{(2)}|\right)$.

[19] Details to come.

5.2 Results

DGP 1. BLP with a Normal Random Coefficient. The first set of simulations use data generated from the benchmark BLP with a normal random coefficient. Table 1 presents the results.

Table 1: DGP 1 – BLP with a Normal Random Coefficient

	Post-Estimation Outputs			Merger's Price Effect		
	Own-Elasticities	Cross-Elasticities	Markups	All firms	Merging Firms	Others
	DGP 1(a): BLP with $\beta \sim \mathcal{N}(3, 6^2)$					
True	-4.0040	0.1479	30.1808	2.7743	6.1053	0.5537
	[-4.0073 ; -4.0007]	[0.1478 ; 0.1481]	[30.1527 ; 30.2089]	[2.7543 ; 2.7943]	[6.0614 ; 6.1491]	[0.5489 ; 0.5584]
BLP	-3.9986	0.1477	30.2406	2.7793	6.1163	0.5546
	[-4.0272 ; -3.9701]	[0.1467 ; 0.1488]	[30.0193 ; 30.4618]	[2.7545 ; 2.8041]	[6.0618 ; 6.1708]	[0.5490 ; 0.5602]
Logit	-2.9766	0.1116	40.7262	3.6330	8.0981	0.6563
	[-3.0455 ; -2.9078]	[0.1090 ; 0.1141]	[39.7618 ; 41.6907]	[3.5473 ; 3.7186]	[7.9074 ; 8.2888]	[0.6406 ; 0.6720]
FIL $D = 4$	-4.2640	0.1626	29.4566	2.8963	6.1789	0.7079
	[-4.4007 ; -4.1273]	[0.1574 ; 0.1679]	[28.4686 ; 30.4447]	[2.8004 ; 2.9922]	[5.9738 ; 6.3840]	[0.6846 ; 0.7312]
FIL $D = 6$	-4.1197	0.1571	30.4607	2.9978	6.3900	0.7364
	[-4.2418 ; -3.9975]	[0.1525 ; 0.1618]	[29.4959 ; 31.4256]	[2.9032 ; 3.0924]	[6.1897 ; 6.5903]	[0.7118 ; 0.7610]
FIL $D = 8$	-4.0918	0.1561	30.7129	3.0328	6.4523	0.7531
	[-4.2157 ; -3.9678]	[0.1514 ; 0.1609]	[29.7299 ; 31.6958]	[2.9353 ; 3.1303]	[6.2469 ; 6.6576]	[0.7271 ; 0.7792]
FIL $D = 10$	-4.1176	0.1572	30.5520	3.0211	6.4223	0.7536
	[-4.2434 ; -3.9917]	[0.1524 ; 0.1620]	[29.5402 ; 31.5638]	[2.9207 ; 3.1214]	[6.2122 ; 6.6325]	[0.7260 ; 0.7811]
	DGP 1(b): BLP with $\beta \sim \mathcal{N}(3, 12^2)$					
True	-3.9814	0.1465	30.3978	2.8406	6.2309	0.5804
	[-3.9849 ; -3.9780]	[0.1464 ; 0.1467]	[30.3679 ; 30.4277]	[2.8127 ; 2.8685]	[6.1693 ; 6.2925]	[0.5737 ; 0.5871]
BLP	-3.9772	0.1464	30.4511	2.8453	6.2413	0.5812
	[-4.0065 ; -3.9479]	[0.1453 ; 0.1475]	[30.2217 ; 30.6804]	[2.8124 ; 2.8782]	[6.1689 ; 6.3138]	[0.5736 ; 0.5889]
Logit	-1.8446	0.0655	67.6284	5.4809	12.4740	0.8189
	[-1.9517 ; -1.7374]	[0.0617 ; 0.0693]	[63.2920 ; 71.9647]	[5.1352 ; 5.8266]	[11.6883 ; 13.2598]	[0.7664 ; 0.8713]
FIL $D = 4$	-4.7170	0.1731	27.0824	2.5132	5.4587	0.5496
	[-4.9548 ; -4.4791]	[0.1644 ; 0.1819]	[25.5676 ; 28.5973]	[2.3719 ; 2.6546]	[5.1544 ; 5.7630]	[0.5166 ; 0.5826]
FIL $D = 6$	-4.4059	0.1614	28.6769	2.6445	5.7584	0.5686
	[-4.5977 ; -4.2141]	[0.1544 ; 0.1685]	[27.3702 ; 29.9836]	[2.5214 ; 2.7675]	[5.4935 ; 6.0233]	[0.5395 ; 0.5976]
FIL $D = 8$	-4.2673	0.1566	30.0540	2.7929	6.0590	0.6155
	[-4.4799 ; -4.0548]	[0.1488 ; 0.1644]	[28.2264 ; 31.8816]	[2.6212 ; 2.9646]	[5.6909 ; 6.4271]	[0.5745 ; 0.6565]
FIL $D = 10$	-4.1669	0.1530	30.7837	2.8644	6.2134	0.6318
	[-4.3754 ; -3.9584]	[0.1454 ; 0.1606]	[28.9413 ; 32.6261]	[2.6889 ; 3.0399]	[5.8380 ; 6.5887]	[0.5893 ; 0.6743]

Notes: Summary statistics across 50 Monte Carlo replications. For each replication, I compute the average. The middle number is the average over replications; lower numbers in brackets are the bounds of the 95% confidence interval.

Consider first DGP 1(a) that allows for moderate unobserved heterogeneity in preferences. Note first that the logit model yields biased estimates of prices elasticities, markups and merger's price effect. This shows that there is sufficient unobserved heterogeneity, so that achieving accuracy is not trivial. Results lead to three main comments. First, they show the ability to my approach to match results from the benchmark BLP. For example, for the own- and cross-price elasticities and the markups, I find that the BLP and FIL confidence intervals contain the true values and overlap. For the merger's price effect, I find a very slight upward bias with respect to BLP and the truth. Overall, I obtain the best results for the post-estimation outputs, when μ is a Bernstein polynomial of order $D = 6$; for the counterfactual analysis (merger's price effect) when it is of order $D = 4$. This suggests that there is a trade-off between in-sample and out-of-sample fits. Second, results illustrate that my approach works well with a sample of moderate size (2500 observations). Third, they show that the FIL confidence intervals are larger than the BLP ones. This was expected since BLP is correctly specified while the FIL model is not. However, my approach still yields informative confidence intervals.

Turn now to the DGP 1(b) that allows for substantial unobserved heterogeneity. Again, the logit model yields (even more) biased results. Results show that my approach is able to match those from the benchmark BLP, even when there is substantial unobserved heterogeneity in preferences.

DGP 2. BLP with a Log-normal Random Coefficient. The second set of simulations use data generated from BLP with a log-normal random coefficient. Figure 2 presents the results.

Table 2: DGP 2 – BLP with a Log-normal Random Coefficient

	Post-Estimation Outputs			Merger's Price Effect		
	Own-Elasticities	Cross-Elasticities	Markups	All firms	Merging Firms	Others
	DGP 2(a): BLP with $\beta \sim 3 + \log \mathcal{N}(0, 3)$					
True - $\log \mathcal{N}$	-4.0015	0.1665	31.149	3.7902	8.3335	0.7613
	[-4.0058 ; -3.9973]	[0.1663 ; 0.1667]	[31.116 ; 31.182]	[3.6837 ; 3.8967]	[8.0746 ; 8.5924]	[0.7518 ; 0.7708]
BLP - $\log \mathcal{N}$	-4.0060	0.1666	31.146	3.7933	8.3417	0.7611
	[-4.0413 ; -3.9707]	[0.1652 ; 0.1681]	[30.865 ; 31.427]	[3.6797 ; 3.9069]	[8.0672 ; 8.6161]	[0.7504 ; 0.7718]
BLP - $\mathcal{N}$	-3.3020	0.1379	37.958	3.9310	8.5451	0.8549
	[-3.3846 ; -3.2195]	[0.1344 ; 0.1413]	[36.917 ; 38.999]	[3.8214 ; 4.0406]	[8.3077 ; 8.7826]	[0.8302 ; 0.8796]
Logit	-1.9854	0.0868	64.566	6.8117	14.917	1.4080
	[-2.0787 ; -1.8921]	[0.0828 ; 0.0909]	[61.355 ; 67.777]	[6.4738 ; 7.1496]	[14.176 ; 15.658]	[1.3385 ; 1.4775]
FIL $D = 4$	-3.9546	0.1710	32.547	3.7111	7.8206	0.9715
	[-4.0710 ; -3.8382]	[0.1660 ; 0.1760]	[31.504 ; 33.590]	[3.5856 ; 3.8366]	[7.5558 ; 8.0854]	[0.9385 ; 1.0045]
FIL $D = 6$	-4.0221	0.1740	31.867	3.6095	7.6272	0.9310
	[-4.1236 ; -3.9206]	[0.1696 ; 0.1783]	[30.951 ; 32.784]	[3.4981 ; 3.7209]	[7.3925 ; 7.8619]	[0.9013 ; 0.9607]
FIL $D = 8$	-3.9870	0.1724	32.174	3.6346	7.6825	0.9360
	[-4.0975 ; -3.8766]	[0.1677 ; 0.1772]	[31.242 ; 33.106]	[3.5253 ; 3.7439]	[7.4501 ; 7.9148]	[0.9082 ; 0.9637]
FIL $D = 10$	-4.0009	0.1730	32.399	3.6567	7.7347	0.9381
	[-4.1349 ; -3.8668]	[0.1672 ; 0.1788]	[30.744 ; 34.055]	[3.4664 ; 3.8470]	[7.3324 ; 8.1369]	[0.8888 ; 0.9874]
FIL $D = 12$	-4.0783	0.1764	31.582	3.5563	7.5249	0.9105
	[-4.2015 ; -3.9552]	[0.1711 ; 0.1817]	[30.334 ; 32.829]	[3.4140 ; 3.6985]	[7.2244 ; 7.8253]	[0.8733 ; 0.9478]
FIL $D = 14$	-4.1339	0.1787	30.883	3.4799	7.3605	0.8928
	[-4.2204 ; -4.0475]	[0.1750 ; 0.1825]	[30.201 ; 31.565]	[3.3959 ; 3.5639]	[7.1827 ; 7.5383]	[0.8709 ; 0.9147]
	DGP 2(b): BLP with $\beta \sim 3 + \log \mathcal{N}(0, 3.5)$					
True - $\log \mathcal{N}$	-3.9973	0.1690	31.378	4.1117	9.0524	0.8179
	[-4.0020 ; -3.9926]	[0.1688 ; 0.1692]	[31.345 ; 31.411]	[3.9616 ; 4.2619]	[8.6863 ; 9.4185]	[0.8053 ; 0.8306]
BLP - $\log \mathcal{N}$	-3.9951	0.1689	31.435	4.1144	9.0571	0.8194
	[-4.0351 ; -3.9551]	[0.1673 ; 0.1706]	[31.115 ; 31.756]	[3.9571 ; 4.2718]	[8.6760 ; 9.4381]	[0.8050 ; 0.8337]
BLP - $\mathcal{N}$	-2.9064	0.1262	43.904	4.7137	10.183	1.0678
	[-3.0127 ; -2.8002]	[0.1216 ; 0.1308]	[42.241 ; 45.567]	[4.5279 ; 4.8995]	[9.7813 ; 10.584]	[1.0253 ; 1.1103]
Logit	-1.5014	0.0681	86.610	9.3526	20.409	1.9817
	[-1.5839 ; -1.4189]	[0.0644 ; 0.0718]	[82.040 ; 91.180]	[8.8532 ; 9.8520]	[19.321 ; 21.497]	[1.8744 ; 2.0889]
FIL $D = 4$	-3.7881	0.1691	34.619	4.2061	8.6338	1.2543
	[-3.9630 ; -3.6132]	[0.1614 ; 0.1768]	[32.855 ; 36.383]	[3.9170 ; 4.4952]	[8.1493 ; 9.1182]	[1.0617 ; 1.4470]
FIL $D = 6$	-3.9206	0.1750	33.361	4.0464	8.2922	1.2159
	[-4.0913 ; -3.7499]	[0.1674 ; 0.1826]	[31.646 ; 35.076]	[3.7315 ; 4.3613]	[7.7922 ; 8.7922]	[0.9920 ; 1.4398]
FIL $D = 8$	-3.8911	0.1737	33.339	4.0134	8.2569	1.1843
	[-4.0391 ; -3.7431]	[0.1671 ; 0.1803]	[31.953 ; 34.726]	[3.7763 ; 4.2504]	[7.8780 ; 8.6359]	[1.0074 ; 1.3612]
FIL $D = 10$	-3.9683	0.1771	32.881	3.9759	8.1591	1.1871
	[-4.1369 ; -3.7997]	[0.1697 ; 0.1846]	[31.247 ; 34.515]	[3.6377 ; 4.3141]	[7.6169 ; 8.7012]	[0.9652 ; 1.4090]
FIL $D = 12$	-4.0025	0.1787	32.151	3.8512	7.9469	1.1208
	[-4.1283 ; -3.8767]	[0.1732 ; 0.1842]	[31.129 ; 33.173]	[3.6337 ; 4.0688]	[7.6130 ; 8.2808]	[0.9583 ; 1.2832]
FIL $D = 14$	-3.9732	0.1774	32.398	3.8917	8.0168	1.1416
	[-4.0993 ; -3.8471]	[0.1718 ; 0.1829]	[31.350 ; 33.447]	[3.6707 ; 4.1127]	[7.6733 ; 8.3604]	[0.9738 ; 1.3094]

Notes: Summary statistics across 50 Monte Carlo replications. For each replication, I compute the average. The middle number is the average over replications; lower numbers in brackets are the bounds of the 95% confidence interval.

Both DGPs 2(a) and 2(b) allow for substantial unobserved heterogeneity in preferences.

In both cases, the logit model yields severely biased results.

This set of simulations aims at comparing my approach to the misspecified BLP that incorrectly assumes a normal, rather than a log-normal, random coefficient. In both DGPs, we observe first that the misspecified BLP yields severely biased estimates of elasticities and markups. This result shows that distributional assumptions in BLP can restrict patterns of substitution and implied markups. See Compiani (2021) and Lu et al. (2019) for other examples. When consider the merger's price effect, the biases almost disappear with DGP 2(a) and are less severe with DGP 2(b). By contrast, my approach provides results very close to the truth, both for the price elasticities and markups and for the merger's price effect. This result shows that my approach can outperform the BLP approach when the latter is misspecified and illustrates how powerful my approach can be by not relying on distributional assumptions and functional form assumptions regarding how substitution patterns depend on product characteristics.

Lastly, I obtain the best results when μ is a Bernstein polynomial of order $D = 12$, against $D = 6$ when the DGPs were generated from the benchmark BLP (DGPs 1(a) and 1(b)). This result is certainly be due to the fact that with the log-normal random coefficients of DGPs 2(a) and 2(b), we have a far higher unobserved heterogeneity in preferences than with the normal random coefficients of DGPs 1(a) and 1(b).

*DGP 3. Gentzkow (2007)'s model of demand for baskets of products***.** The third set of simulations use data generated from Gentzkow (2007). Figure 3 presents the results.

Table 3: DGP 3 – Gentzkow (2007)'s model of demand for baskets of products

	Post-Estimation Outputs			Merger's Price Effect		
	Own-Elasticities	Cross-Elasticities	Markups	All firms	Merging Firms	Others
	DGP: Gentzkow's Model					
True - Baskets	-4.0112	0.1676	31.722	3.3917	7.1085	0.9138
	[-4.0138 ; -4.0085]	[0.1675 ; 0.1677]	[31.694 ; 31.75]	[3.3720 ; 3.4113]	[7.0664 ; 7.1506]	[0.9087 ; 0.9188]
BLP	-4.4015	0.1815	27.917	2.8976	6.2044	0.6931
	[-4.4546 ; -4.3484]	[0.1793 ; 0.1836]	[27.585 ; 28.249]	[2.8612 ; 2.9340]	[6.1261 ; 6.2827]	[0.6843 ; 0.7019]
Logit	-3.3308	0.1392	36.851	3.7974	8.1757	0.8786
	[-3.3899 ; -3.2716]	[0.1367 ; 0.1416]	[36.186 ; 37.517]	[3.7282 ; 3.8667]	[8.0267 ; 8.3246]	[0.8624 ; 0.8948]
FIL $D = 6$	-3.8496	0.1604	32.721	3.5872	7.368	1.0667
	[-3.9700 ; -3.7293]	[0.1553 ; 0.1654]	[31.611 ; 33.832]	[3.4617 ; 3.7127]	[7.1131 ; 7.6228]	[1.0271 ; 1.1063]
FIL $D = 8$	-3.8844	0.1618	32.331	3.5411	7.2782	1.0497
	[-3.9931 ; -3.7758]	[0.1573 ; 0.1664]	[31.372 ; 33.291]	[3.4338 ; 3.6485]	[7.0601 ; 7.4964]	[1.0160 ; 1.0834]
FIL $D = 10$	-3.8773	0.1615	32.425	3.5511	7.2988	1.0527
	[-3.9902 ; -3.7643]	[0.1568 ; 0.1662]	[31.421 ; 33.429]	[3.4395 ; 3.6628]	[7.0711 ; 7.5266]	[1.0181 ; 1.0873]
FIL $D = 12$	-3.8971	0.1623	32.33	3.5422	7.2773	1.0521
	[-4.0178 ; -3.7763]	[0.1573 ; 0.1674]	[31.227 ; 33.433]	[3.4192 ; 3.6651]	[7.0260 ; 7.5285]	[1.0145 ; 1.0897]
FIL $D = 14$	-3.9127	0.163	32.113	3.5168	7.2287	1.0423
	[-4.0237 ; -3.8018]	[0.1583 ; 0.1676]	[31.148 ; 33.078]	[3.4092 ; 3.6244]	[7.0097 ; 7.4477]	[1.0087 ; 1.0758]
FIL $D = 16$	-3.8884	0.162	32.113	3.5452	7.2873	1.0504
	[-4.0058 ; -3.7711]	[0.1571 ; 0.1669]	[31.148 ; 33.078]	[3.4266 ; 3.6638]	[7.0449 ; 7.5298]	[1.0142 ; 1.0866]

Notes: Summary statistics across 50 Monte Carlo replications. For each replication, I compute the average. The middle number is the average over replications; lower numbers in brackets are the bounds of the 95% confidence interval.

Results show that the benchmark BLP can lead to biased estimates of substitution patterns, markups and merger price effects when consumers choose baskets of products, rather than products alone. This result is theoretically shown by Narita and Saito (2021). By constrast, my approach obtains good predictions of substitution patterns, markups and merger price effects, especially when $D = 14$. This demonstrates that my approach can outperform BLP when the DGP is generated by consumers making multiple choices.

DGP 4. FIL Model with a Linear Mapping. The third set of simulations use data generated from a FIL model with a linear mapping. Figure **??** presents the results.

Table 4: DGP 4 – FIL Model with a Linear Mapping

	Post-Estimation Outputs			Merger's Price Effect		
	Own-Elasticities	Cross-Elasticities	Markups	All firms	Merging Firms	Others
	DGP 4(a): FIL with $\mu_{ij} = 0.04(1 - \lvert x_{it}^{(2)} - x_{jt}^{(2)}\rvert)$					
FIL	-2.2866	0.0864	52.371	4.9727	10.15	1.5213
[-2.2875 ; -2.2857]	[0.0864 ; 0.0865]	[52.35 ; 52.391]	[4.9606 ; 4.9849]	[10.125 ; 10.175]	[1.5171 ; 1.5255]	
BLP	-1.7152	0.0612	70.136	5.5674	12.557	0.9079
[-1.7862 ; -1.6441]	[0.0587 ; 0.0638]	[66.939 ; 73.333]	[5.3155 ; 5.8193]	[11.986 ; 13.127]	[0.8681 ; 0.9476]	
	DGP 4 (b): FIL with $\mu_{ij} = 0.02 + 0.02(1 - \lvert x_{it}^{(2)} - x_{jt}^{(2)}\rvert)$					
FIL	-4.0692	0.1416	32.571	4.9467	7.185	3.4545
	[-4.0802 ; -4.0582]	[0.1414 ; 0.1418]	[32.554 ; 32.589]	[4.8787 ; 5.0147]	[7.1026 ; 7.2673]	[3.3435 ; 3.5655]
BLP	-5.1602	0.1679	22.969	1.6898	3.8511	0.2489
	[-5.2555 ; -5.0649]	[0.1648 ; 0.171]	[22.547 ; 23.39]	[1.6585 ; 1.7211]	[3.7793 ; 3.9229]	[0.2444 ; 0.2534]

Notes: Summary statistics across 50 Monte Carlo replications. For each replication, I compute the average. The middle number is the average over replications; lower numbers in brackets are the bounds of the 95% confidence interval.

Results show that the benchmark BLP (with normal random coefficient) can lead to incorrect estimates of substitution patterns, markups and merger price effects. These simulations thus illustrate that, even when it allows only substitutable products, my approach can generate patterns of substitution, markups and merger price effect that the benchmark BLP cannot recover.

6 Conclusion

This paper has proposed a flexible approach to estimate substitution patterns in differentiated products markets using market-level data. My approach relies on an inverse market share function rather than a utility function. Thereby, even if it prevents studies at the individual consumer level, it can be used to answer many economic questions of interest, such as the measurement of market power and effects of regulatory changes in taxes and trade policy. Furthermore, my approach does not rely on arbitrary restrictions on how substitution patterns depend on product characteristics and is agnostic about individual consumer behavior. This explains why it can accommodate rich substitution patterns, including some that the BLP approach cannot recover.

My approach can be applied to various topics in industrial organization, international trade, digital economics, environmental economics, etc. In particular, due to its simplicity of estimation, a likely audience of my approach involves antitrust practitioners who are under time pressure and wish to avoid complex estimation procedures without sacrificing flexibility. Indeed, the nested logit models are commonly used by antitrust practitioners for merger simulation.[20] Yet, they have been criticized for restricting the substitution patterns and they require the modeler to select the relevant observed characteristics that define groups, which, in practice, is not always obvious.[21] By contrast, my approach does not require a priori restrictions, such as product segmentation, which makes it a good option for merger simulation purposes.

Throughout this paper, I have maintained several assumptions that may be relaxed. First, my demand model generates market shares (unit demand). An extension to variable consumption, in spirit to the constant expenditure specifications of Björnerstedt and Verboven (2016), will be considered in a future version. Second, I have assumed that the modeler has only access to market-level data. An extension to individual-level data is left for future research. Lastly, a future version of this paper will include an empirical application and an extension to multi-dimensional mapping.

Appendix A Bernstein Polynomials

This appendix provides some details on the Bernstein polynomials. See Chapter 6 in Davis (1975) for a comprehensive treatment.

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

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

[20]See e.g. the Lagardère/Natexis/VUP (2004), TomTom/Tele Atlas (2008), Unilever/Sara Lee (2010) cases litigated by the European Commission (CCR - Competition Competence Report Autumn 2013/1). Despite their limitations, they are also used by academics. (See e.g., Björnerstedt and Verboven, 2016; Berry et al., 2016, for recent papers).

[21]Consider, for example, the market for cars, where cars belong to five segments: subcompact, compact, standard, intermediate, and luxury. Grigolon (2021) suggests a natural ordering of cars from subcompact to luxury, while Brenkers and Verboven (2006) consider a hierarchical structure without prior ordering. Determining which of the two nesting structures best describes the market is not obvious.

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

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

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

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

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

The generalization to a multivariate function defined over $[a_1, b_1] \times \ldots \times [a_L, b_L]$ is straightforward. A multivariate function can be approximated by

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

for $(x_1, \ldots, x_L) \in [a_1, b_1] \times \ldots \times [a_L, b_L]$ and for some coefficients $\theta_{k_1,\ldots,k_L}$, $k_1, k_2, \ldots = 0, \ldots, D$.

Appendix B Elements of Convex Analysis

This appendix provides the elements of convex analysis used in the paper. See Rockafellar (1970) for a comprehensive treatment of the topic. For a a real-valued function f, the vector let $\nabla f(\mathbf{x}) = [\partial f(\mathbf{x})/\partial x_i]$ denote its gradient with respect to the vector $\mathbf{x}$. For a set $\mathcal{X}$, $\mathrm{int}(\mathcal{X})$ its interior and $\mathrm{bd}(\mathcal{X})$ its boundary.

Consider a convex function $f : \mathcal{X} \subseteq \mathbb{R}^J \to \mathbb{R} \cup \{\pm\infty\}$. Its *effective domain* $\mathrm{dom} f$ is defined by $\mathrm{dom} f = \{\mathbf{x} \in \mathcal{X} | f(\mathbf{x}) < +\infty\}$ and is a convex set in $\mathbb{R}^J$. A *proper convex function* f is a convex function that takes values in the extended real number line such that $f(\mathbf{x}) < +\infty$ for at least one $\mathbf{x}$ and $f(\mathbf{x}) > -\infty$ for every $\mathbf{x}$. Then, f is proper if and only if its effective domain $\mathrm{dom} f$ is non-empty and the restriction of f to $\mathrm{dom} f$ is finite.

Let $f : \mathbb{R}^J \to \mathbb{R} \cup \{+\infty\}$. Its *convex conjugate* is the function $f^* : \mathbb{R}^J \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(\mathbf{x}^*) = \sup_{\mathbf{x} \in \mathrm{dom} f} \{\mathbf{x}^{*\intercal}\mathbf{x} - f(\mathbf{x})\}.$$

A proper convex function f is *essentially smooth* if (i) $\operatorname{int}(\operatorname{dom} f)$ is non-empty; (ii) f is differentiable throughout $\operatorname{int}(\operatorname{dom} f)$, and (iii) $\lim_{i\to\infty} |\nabla f(\mathbf{x}_i)| = +\infty$ whenever $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is a sequence in $\operatorname{int}(\operatorname{dom} f)$ converging to a point $\mathbf{x} \in \operatorname{bd}(\operatorname{dom} f)$.

A pair $(\operatorname{int}(\operatorname{dom} f), f)$ is a *convex function of Legendre type* if $\operatorname{int}(\operatorname{dom} f)$ is an open convex set and f is a strictly convex function on $\operatorname{int}(\operatorname{dom} f)$ that is essentially smooth.

The following proposition, due to Rockafellar (1970), establishes an important invertibility result.

Proposition 1. Let $f : \mathbb{R}^J \to \mathbb{R} \cup \{+\infty\}$ be a continuous convex function. Assume that $(\operatorname{int}(\operatorname{dom} f), f)$ is a convex function of Legendre type. Then, ∇f is a continuous bijection between $\operatorname{int}(\operatorname{dom} f)$ and $\operatorname{int}(\operatorname{dom} f^*)$, with a continuous inverse mapping $(\nabla f)^{-1} = \nabla f^*$, i.e., $\nabla f^*(\mathbf{x}^*) = (\nabla f)^{-1}(\mathbf{x}^*)$ for all $\mathbf{x}^* \in \operatorname{int}(\operatorname{dom} f^*)$.

Proof. See Theorem 26.5 in Rockafellar (1970). $\square$

Appendix C Proofs

C.1 Invertibility

Let $\Delta_J \equiv \left\{\mathbf{s}_t \in [0,\infty)^J : \sum_{j=1}^J s_{jt} < 1\right\}$ with $\operatorname{int}(\Delta_J)$ its (relative) interior and $\operatorname{bd}(\Delta_J)$ its boundary. The following proposition establishes an invertibility result that can be used to show invertibility of a large class of inverse demand functions.

Proposition 2. Consider the function $\mathbf{G} = (G_1, \ldots, G_J) : [0,\infty)^J \to [0,\infty)^J$. Assume that $\mathbf{G}$ is continuously differentiable and homogeneous of degree one on $\operatorname{int}(\Delta_J)$ and that $\ln \mathbf{G}$ has a matrix of derivatives $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}$ that is positive definite and symmetric on $\operatorname{int}(\Delta_J)$. Further assume that the 1-norm $|\ln \mathbf{G}(\mathbf{s}_t)|$ approaches infinity as $\mathbf{s}_t$ approaches $\operatorname{bd}(\Delta_J)$. Let $\mathbf{f} : [0,\infty)^J \to \mathbb{R}^J$ be defined by $\mathbf{f} = (f_1, \ldots, f_J)$ where $f_j(\mathbf{s}_t) = \ln G_j(\mathbf{s}_t) - \ln\left(1 - \sum_{k=1}^J s_{kt}\right)$. It follows that $\mathbf{f}$ is invertible on $\operatorname{int}(\Delta_J)$.

Proof. The proof is an application of Proposition 1 to the pair $(\operatorname{int}(\Delta_J), \Omega)$, where Ω is defined by

$$\Omega(\mathbf{s}_t) = \begin{cases} \sum_{j=1}^J s_{jt} f_j(\mathbf{s}_t) + \ln\left(1 - \sum_{k=1}^J s_{kt}\right) - \sum_{j=1}^J s_{jt} & \text{if } \mathbf{s}_t \in \Delta_J, \\ +\infty & \text{otherwise.} \end{cases}$$

Using the generalized Euler equation for homothetic functions (McElroy, 1969), we have $\nabla\Omega(\mathbf{s}_t) = \mathbf{f}(\mathbf{s}_t)$. The proof thus consists in showing that the pair $(\operatorname{int}(\Delta_J), \Omega)$ is a convex function of Legendre type.

Ω is strictly convex on $\operatorname{int}(\Delta_J)$, since its Hessian is equal to $\mathbf{J}^{\mathbf{s}}_{\ln \mathbf{G}}(\mathbf{s}_t) + \mathbf{1}_{JJ}/s_{0t}$ for any $\mathbf{s} \in \operatorname{int}(\Delta_J)$, where $\mathbf{1}_{JJ}$ is a $J \times J$ matrix of ones. Ω is essentially smooth, since it is differentiable through the open convex set $\operatorname{int}(\Delta_J)$ with $\lim_{i\to\infty} |\nabla\Omega\left(\mathbf{s}^i_t\right)| = +\infty$ whenever $\mathbf{s}^1_t, \mathbf{s}^2_t, \ldots$ is a sequence in $\operatorname{int}(\Delta_J)$ converging to a point $\mathbf{s}_t \in \operatorname{bd}(\Delta_J)$. This latter feature is shown by first noting that $\nabla\Omega(\mathbf{s}_t) = \mathbf{f}(\mathbf{s}_t)$ for $\mathbf{s}_t \in \operatorname{int}(\Delta_J)$ and then using that $\lim_{\mathbf{s}_t\to\operatorname{bd}(\Delta_J)} |\ln \mathbf{G}(\mathbf{s}_t)| = +\infty$. □

As a corollary, I obtain invertibility of the FIL model.

Corollary 1. Let Restrictions (R1) – (R3) hold. Consider any vector $\boldsymbol{\delta}_t \in \mathbb{R}^J$ of product indexes. Then, holding $\mathbf{x}^{(2)}_t$ fixed and $\delta_{0t} = 0$, there exists a unique vector $\mathbf{s}_t \in \Delta^+_J$ of nonzero market shares such that Equations (2) hold.

Equations (2) describe an inverse demand function $\boldsymbol{\sigma}^{-1}$, i.e., a mapping from market shares to product indexes. This corollary establishes existence and uniqueness of the inverse mapping from product indexes to market shares, i.e., the demand function $\boldsymbol{\sigma}$, up to the normalization $\delta_{0t} = 0$.

Proof. The proof amounts to show that the FIL model satisfies the assumptions of Proposition 2. Let

$$G_j(\mathbf{s}_t) = (s_{jt})^{1-\sum_{i\neq j}\mu_{ij}} \prod_{i\neq j} (s_{it} + s_{jt})^{\mu_{ij}} \tag{32}$$

then $f_j = \ln G_j$ corresponds to the FIL model (2). The matrix of derivatives of $\ln \mathbf{G}$ has entries ij given by

$$\left[\frac{1-\sum_{i\neq j}\mu_{ij}}{s_{jt}} + \sum_{i\neq j}\frac{\mu_{ij}}{s_{it}+s_{jt}}\right]\mathbf{1}\{i=j\} + \left[\frac{\mu_{ij}}{s_{it}+s_{jt}}\right]\mathbf{1}\{i\neq j\}.$$

It is easy to show that the 1-norm $|\ln \mathbf{G}\left(\mathbf{s}_t\right)|$ that approaches infinity as $\mathbf{s}_t$ approaches $\operatorname{bd}\left(\Delta_J\right)$. It then remains to show that the function $\mathbf{G}$ is homogeneous of degree one and that its derivative matrix is positive definite and symmetric.

It is homogeneous of degree one, since for $\kappa > 0$ and $j = 1, \ldots, J$,

$$\begin{aligned} G_j(\kappa \mathbf{s}_t) &= (\kappa s_{jt})^{1-\sum_{i \neq j} \mu_{ij}} \prod_{i \neq j} [\kappa (s_{it} + s_{jt})]^{\mu_{ij}} \\ &= \left[\kappa^{1-\sum_{i \neq j} \mu_{ij}} \prod_{i \neq j} \kappa^{\mu_{ij}}\right] \left[(s_{jt})^{1-\sum_{i \neq j} \mu_{ij}} \prod_{i \neq j} (s_{it} + s_{jt})^{\mu_{ij}}\right], \\ &= \left[\kappa^{1-\sum_{i \neq j} \mu_{ij} + \sum_{i \neq j} \mu_{ij}}\right] G_j(\mathbf{s}_t), \\ &= \kappa G_j(\mathbf{s}_t). \end{aligned}$$

Furthermore, the derivative matrix of $\ln \mathbf{G}$ is symmetric since Restriction (R1) implies that its entry ij, $\mu_{ij}/(s_i + s_j)$, equals its entry ji, $\mu_{ji}/(s_j + s_i)$.

Lastly, the derivative matrix of $\ln \mathbf{G}$ is positive definite as, by Restrictions (R2) and (R3), it is a symmetric, strictly diagonally dominant matrix with positive diagonal entries (Horn and Johnson, 2012, Theorem 6.1.10.). □

C.2 Flexibility

The following proposition establishes that the FIL model is flexible in the sense of Diewert (1974) in a large class of inverse demand models.

Proposition 3. The FIL model is flexible in the sense of Diewert (1974) in the class of inverse demand models defined by

$$\sigma_j^{-1}(\mathbf{s}_t; \boldsymbol{\mu}) = \ln G_j(\mathbf{s}_t; \boldsymbol{\mu}) - \ln(s_{0t}) \tag{33}$$

where

1. The function $\mathbf{G}$ is homogeneous of degree one,
2. The function $\ln \mathbf{G}$ has a matrix of derivatives that is symmetric and positive definite on Δ_J,
3. The 1-norm $|\ln \mathbf{G}(\mathbf{s}_t)$ approaches infinity as $\mathbf{s}_t$ approaches $\mathrm{bd}(\Delta_J)$.

Proof. Assume that we observe the vectors of prices and market shares, $\mathbf{p}_t$ and $\mathbf{s}_t$. The proof consists in showing that the FIL model satisfies the following two requirements: (i) it can match the vector of market shares $\mathbf{s}_t$; and (ii) it can match the true matrix of own-and cross-price demand elasticities, with entries ij given by λ_{ij}.

Market Shares. Proposition 1 implies that there exists a unique vector of $\boldsymbol{\delta}_t$ such that $\boldsymbol{\sigma}(\boldsymbol{\delta}_t; \boldsymbol{\mu}) = \mathbf{s}_t$, which shows the first requirement. Given a value for $\boldsymbol{\mu}$, we can set

$$\delta_{jt} = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sum_{i \neq j} \mu_{ij} \ln\left(\frac{s_{jt}}{s_{it} + s_{jt}}\right), \quad j = 1, \ldots, J.$$

to match the vector of market shares $\mathbf{s}_t$.

Price Elasticities. Observing prices and market shares, matching price elasticities amounts to matching price derivatives. Since the matrix of price elasticities is positive definite, this is equivalent to matching its inverse whose entries ij are given by

$$\lambda_{ij} \equiv \frac{1}{\alpha} \frac{\partial \sigma_i^{-1}(\mathbf{s}_t)}{\partial s_{jt}} = \begin{cases} \dfrac{1}{\alpha}\left[\dfrac{1}{s_{0t}} + \dfrac{1 - \sum_{i \neq j} \mu_{ij}}{s_{jt}} + \sum_{i \neq j} \dfrac{\mu_{ij}}{s_{it} + s_{jt}}\right], & \text{if } \ i = j, \\ \\ \dfrac{1}{\alpha}\left[\dfrac{1}{s_{0t}} + \dfrac{\mu_{ij}}{s_{it} + s_{jt}}\right], & \text{if } \ i \neq j. \end{cases} \tag{34}$$

Consider matching the off-diagonal entry λ_{ij}. Given a value for α, we must choose μ_{ij} such that

$$\lambda_{ij} = \frac{1}{\alpha}\left(\frac{1}{s_{0t}} + \frac{\mu_{ij}}{s_{it} + s_{jt}}\right), \tag{35}$$

which can be inverted to obtain μ_{ij} as a function of $\mathbf{s}_t$, λ_{ij} and α

$$\mu_{ij} = \left(\alpha \lambda_{ij} - \frac{1}{s_{0t}}\right)(s_{it} + s_{jt}). \tag{36}$$

Equation (35), together with Restrictions (R2), we note that $\lambda_{ij} > 0$. Note that we must choose sufficiently high for Restrictions (R2) to be satisfied and sufficiently for Restrictions (R3) to be satisfied.

Consider now matching the diagonal entries λ_{jj}. Differentiating $\sum_{k=1}^{J} \sigma_k(\boldsymbol{\delta}) + \sigma_0(\boldsymbol{\delta}) = 1$ implies, for all $j = 1, \ldots, J$, that $\sum_{k=1}^{J} \frac{\partial \sigma_k(\boldsymbol{\delta})}{\partial p_j} + \frac{\partial \sigma_0(\boldsymbol{\delta})}{\partial p_j} = 0$, which can be rearranged as $\sum_{k=1}^{J} \frac{\partial \sigma_k(\boldsymbol{\delta})}{\partial p_j} = -\alpha s_0 s_j$. This rearrangement uses that the relationship between the outside good and any product $j > 0$ is of the logit form and that market shares are matched.

C.3 Mapping

Consider the mappings (3)

$$\mu_{ij} = \mu\left(\mathbf{d}^{(2)}_{ij,t}\left(\mathbf{x}^{(2)}_{it}, \mathbf{x}^{(2)}_{jt}\right)\right) \tag{37}$$

where I make explicit the dependency of $\mathbf{d}^{(2)}_{ij,t}$ in characteristics $\mathbf{x}^{(2)}_{jt}$ for the purpose of the proof.

The FIL model (2) mapped in characteristic space can be rewritten as

$$\sigma_j^{-1}\left(\mathbf{s}_t, \mathbf{x}^{(2)}_t\right) = \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sum_{i \neq j} \mu\left(\mathbf{d}^{(2)}_{ij,t}\left(\mathbf{x}^{(2)}_{it}, \mathbf{x}^{(2)}_{jt}\right)\right) \ln\left(\frac{s_j}{s_{it} + s_{jt}}\right), \tag{38}$$

$$= \ln\left(\frac{s_{jt}}{s_{0t}}\right) - \sum_{i=1}^{J} \mu\left(\mathbf{d}^{(2)}_{ij,t}\left(\mathbf{x}^{(2)}_{it}, \mathbf{x}^{(2)}_{jt}\right)\right) \ln\left(\frac{s_j}{s_{it} + s_{jt}}\right) + C_t, \tag{39}$$

where $C_t = \mu(\mathbf{0}) \ln(1/2) \in \mathbb{R}$.

Observe in Equation (39) that the sum is over all products $i = 1, \ldots, J$, including product j itself, and that the constant C is product-invariant. This shows that the FIL model projected in characteristics space is symmetric and anomynous.

Lastly, using Equation (38) shows that σ_j^{-1} is invariant to translation in $\mathbf{x}^{(2)}$ since, for all $c \in \mathbb{R}$,

$$\sigma_j^{-1}\left(\mathbf{s}, \mathbf{x}^{(2)} + c\mathbf{1}\right) = \sigma_j^{-1}\left(\mathbf{s}, \mathbf{d}^{(2)}\left(\mathbf{x}^{(2)} + c\mathbf{1}\right)\right) = \sigma_j^{-1}\left(\mathbf{s}, \mathbf{d}^{(2)}\left(\mathbf{x}^{(2)}\right)\right) = \sigma_j^{-1}\left(\mathbf{s}, \mathbf{x}^{(2)}\right).$$

References

ALLEN, R. AND J. REHBECK (2019): "Identification With Additively Separable Heterogeneity," *Econometrica*, 87, 1021–1054.

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

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

ATHEY, S. AND G. W. IMBENS (2007): "Discrete choice models with multiple unobserved choice characteristics," *International Economic Review*, 48, 1159–1192.

BAJARI, P. AND C. L. BENKARD (2005): "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach," *Journal of political economy*, 113, 1239–1276.

BAKER, J. B. AND T. F. BRESNAHAN (1985): "The gains from merger or collusion in product-differentiated industries," *The Journal of Industrial Economics*, 427–444.

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

BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80, 2369–2429.

BERRY, S., A. EIZENBERG, AND J. WALDFOGEL (2016): "Optimal product variety in radio markets," *The RAND Journal of Economics*, 47, 463–497.

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

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

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

——— (1999): "Voluntary export restraints on automobiles: Evaluating a trade policy," *American Economic Review*, 89, 400–430.

BERRY, S. AND A. PAKES (2007): "The pure characteristics demand model," *International Economic Review*, 48, 1193–1225.

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

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

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

BOURREAU, M., Y. SUN, AND F. VERBOVEN (2021): "Market Entry, Fighting Brands and Tacit Collusion: Evidence from the French Mobile Telecommunications Market," .

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

BRESNAHAN, T. F. (1989): "Empirical studies of industries with market power," *Handbook of industrial organization*, 2, 1011–1057.

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

CHAK, P. M., N. MADRAS, AND B. SMITH (2005): "Semi-nonparametric estimation with Bernstein polynomials," *Economics Letters*, 89, 153–156.

CHAMBERLAIN, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

CHETVERIKOV, D., A. SANTOS, AND A. M. SHAIKH (2018): "The econometrics of shape restrictions," *Annual Review of Economics*, 10, 31–63.

CHU, C. (1989): "A Paired Combinational Logit Model for Travel Demand Analysis," in *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, Vol. 4, Western Periodicals, Ventura, CA, 295–309.

COMPIANI, G. (2021): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," *Unpublished, available on author's website*.

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

CRAWFORD, G. S. AND A. YURUKOGLU (2012): "The welfare effects of bundling in multichannel television markets," *American Economic Review*, 102, 643–85.

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

DAVIS, P. J. (1975): *Interpolation and approximation*, Courier Corporation.

DEATON, A. AND J. MUELLBAUER (1980): "An almost ideal demand system," *The American economic review*, 70, 312–326.

DICK, A. A. (2008): "Demand estimation and consumer welfare in the banking industry," *Journal of Banking & Finance*, 32, 1661–1676.

DIEWERT, W. (1974): "Application of Duality Theory," *Frontiers of Quantitative Economics*, 2.

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.

FOSGERAU, M., J. MONARDO, AND A. DE PALMA (2021): "The Inverse Product Differentiation Logit Model," *Available at SSRN: https://ssrn.com/abstract=3141041*.

FOX, J. T. AND A. GANDHI (2016): "Nonparametric identification and estimation of random coefficients in multinomial choice models," *The RAND Journal of Economics*, 47, 118–139.

FOX, J. T., K. IL KIM, AND C. YANG (2016): "A simple nonparametric approach to estimating the distribution of random coefficients in structural models," *Journal of Econometrics*, 195, 236–254.

FOX, J. T., K. I. KIM, S. P. RYAN, AND P. BAJARI (2011): "A simple estimator for the distribution of random coefficients," *Quantitative Economics*, 2, 381–418.

GANDHI, A. AND J.-F. HOUDE (2020): "Measuring substitution patterns in differentiated products industries," *Available at SSRN: https://ssrn.com/abstract=3472810*.

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

GORDON, B. R. AND W. R. HARTMANN (2013): "Advertising effects in presidential elections," *Marketing Science*, 32, 19–35.

GRIFFITH, R., M. O'CONNELL, AND K. SMITH (2019): "Tax design in the alcohol market," *Journal of Public Economics*, 172, 20–35.

GRIGOLON, L. (2021): "Blurred boundaries: a flexible approach for segmentation applied to the car market," *Forthcoming at Quantitative Economics*.

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

HO, K. AND J. ISHII (2011): "Location and competition in retail banking," *International Journal of Industrial Organization*, 29, 537–546.

HO, K. AND R. S. LEE (2017): "Insurer competition in health care markets," *Econometrica*, 85, 379–417.

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

HURWICZ, L. (1966): "On the structural form of interdependent systems," in *Studies in Logic and the Foundations of Mathematics*, Elsevier, vol. 44, 232–239.

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

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

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

LU, Z., X. SHI, AND J. TAO (2019): "Semi-Nonparametric Estimation of Random Coefficient Logit Model for Aggregate Demand," *Available at SSRN 3503560*.

MATZKIN, R. L. (1994): "Restrictions of economic theory in nonparametric methods," *Handbook of econometrics*, 4, 2523–2558.

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

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

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 15, 447–470.

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

Narita, Y. and K. Saito (2021): "Approximating Choice Data by Discrete Choice Models," *Unpublished, available on author's website*.

Nevo, A. (2000a): "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *The RAND Journal of Economics*, 31, 395–421.

——— (2000b): "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of economics & management strategy*, 9, 513–548.

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

Newey, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica: Journal of the Econometric Society*, 58, 809–837.

Nocke, V. and N. Schutz (2017): "Quasi-linear integrability," *Journal of Economic Theory*, 169, 603–628.

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

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

Pinkse, J., M. E. Slade, and C. Brett (2002): "Spatial price competition: a semiparametric approach," *Econometrica*, 70, 1111–1153.

REYNAERT, M. (2021): “Abatement strategies and the cost of environmental regulation: Emission standards on the European car market,” *The Review of Economic Studies*, 88, 454–488.

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

SALANIÉ, B. AND F. A. WOLAK (2019): “Fast," robust", and approximately correct: estimating mixed demand systems,” *National Bureau of Economic Research*.

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

SIMON, C. P. AND L. BLUME (1994): *Mathematics for economists*, New York, NY: W.W. Norton Company.

SLADE, M. E. (2004): “Market power and joint dominance in UK brewing,” *The Journal of Industrial Economics*, 52, 133–163.

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

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

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