WARWICK

ECONOMICS

A BLP Demand Model of Product-Level Market Shares with Complementarity

Ao Wang

April 2021 No: 1351

Warwick Economics Research Papers

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

A BLP Demand Model of Product-Level Market Shares with Complementarity

Ao Wang*

January 2021

Abstract

Applied researchers most often estimate the demand for differentiated products assuming that at most one item can be purchased. Yet simultaneous multiple purchases are pervasive. Ignoring the interdependence among multiple purchases can lead to erroneous counterfactuals, in particular, because complementarities are ruled out. I consider the identification and estimation of a random coefficient discrete choice model of bundles, namely sets of products, when only product-level market shares are available. This last feature arises when only aggregate purchases of products, as opposed to individual purchases of bundles, are available, a very common phenomenon in practice. Following the classical approach with aggregate data, I consider a two-step method. First, using a novel inversion result in which demand can exhibit Hicksian complementarity, I recover the mean utilities of products from product-level market shares. Second, to infer the structural parameters from the mean utilities while dealing with price endogeneity, I use instrumental variables. I propose a practically useful GMM estimator whose implementation is straightforward, essentially as a standard BLP estimator. Finally, I estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. The demand estimates suggest that RTE cereals and milk are overall complementary and the synergy in consumption crucially depends on their characteristics. Ignoring such complementarities results in misleading counterfactuals.

*University of Warwick, UK (ao.wang@warwick.ac.uk). I am indebted to Xavier D'Haultfœuille and Alessandro Iaria for the fruitful discussions and continued encouragement at each stage of this paper. I am grateful to Jaap Abbring, Philippe Choné, Laurent Lamy, Roxana Fernández Machado, Isabelle Méjean, Eric Renault, and Matt Shum for their insightful suggestions. I would like to thank seminar and conference participants at CREST, University of Surrey, University of Warwick, HEC Montréal, University of Western Ontario, VIOS China, SEA 90th Annual Meeting. I would also like to thank IRI for making the data used in this paper available. All estimates and analyses in this paper, based on data provided by IRI, are by mine and not by IRI. All errors are my own.

1 Introduction

Since the seminal work of Berry (1994) and Berry et al. (1995) (henceforth BLP), BLP-type models have been widely used in empirical literature of demand estimation.[1] Researchers most often estimate models of demand for single products, assuming that individual purchases at most one single product. Yet, the behaviour of making multiple simultaneous purchases is pervasive. In many economic analyses (e.g. multi-category demand, nonlinear pricing, supermarket competition), interdependence among multiple purchases in demand is key, which is, however, assumed away in models of demand for single products. In particular, Hicksian complementarities are ruled out.[2] As a result, estimating demand models of single products may lead to biased estimates and misleading policy counterfactuals.[3]

This paper proposes a random coefficient discrete-choice model of demand for bundles.[4] This model has various notable advantages. First, its application only requires the availability of aggregate demand data at product level (e.g. aggregate choice probability, or sales quantities, of products) defined in form of "market shares". Such data is widely accessible in most industries.[5] Differently, models of demand for bundles routinely used in the empirical literature often rely on the availability of individual-level choice data of bundles (e.g. individual transaction data), which may be costly to obtain. Second, different from models of demand for single products, the proposed model does not restrict products to be substitutes. It incorporates individuals' behaviour of multiple purchases, allowing for Hicksian complementarities among products. Notably, the model enables to encompass various mechanisms that can drive Hicksian complementarity, while still allowing for endogenous prices.[6] Third, the identification arguments are constructive and lead to a practically useful Generalized Method of Moments (GMM) estimator. In particular, it can handle potentially large choice sets in multiple purchases. Finally, the implementation of the GMM estimator is straightforward, essentially as a standard BLP estimator.

To motivate the identification discussion, I distinguish two sets of demand prim-

[1]BLP-type models also gain popularity outside of empirical industrial organisation, e.g. analysis of voting data (Rekkas (2007), Milligan and Rekkas (2008), Gordon and Hartmann (2013), Merlo and Paula (2017), Gillen et al. (2019)), asset pricing (Koijen and Yogo, 2019).

[2]Hicksian complementarity is defined as negative (compensated) cross-price elasticity between two products. For a survey of different concepts of complementarity, see Samuelson (1974).

[3]See section 3.3 for Monte Carlo evidences of such bias in various counterfactual analyses.

[4]In the empirical literature, the terminology "bundle" is often referred to as a set of products purchased by individuals. In this paper, I use this terminology and formalise it in Assumption 1.

[5]This data requirement is the same as classical BLP models of single products.

[6]Examples of such mechanisms include shopping cost (Pozzi (2012), Thomassen et al. (2017)), preference for variety (Hendel (1999), Dubé (2004)), and synergies in consumption (Gentzkow, 2007).

itives that are sufficient for different kinds of economic analyses in models of demand for bundles. The first set includes product-level market share functions; I show that they are sufficient for economic analyses (e.g. identification of price elasticities, marginal costs, and mergers) under linear pricing.[7] The second set includes bundle-level market share functions; I prove that they are sufficient for the analyses under nonlinear pricing.[8]

I then organise the identification discussion in two parts. In the first part, as in classical demand models using aggregate data, I use a two-step strategy to identify product-level market share functions. In the first step, I invert the observed product-level market shares to the mean utilities of products. Because of possible Hicksian complementarities among products, the typical conditions that guarantee the invertibility of product-level market share functions (connected substitutes conditions, see Berry et al. (2013)) may not hold. To solve this challenge, I use a novel demand inverse argument that hinges on two elements. First, the affine relationship between the utilities of bundles and its single products: the average utility of any bundle equals the sum of those of its single products plus an extra term capturing their potential *demand synergy*. Second, the P-matrix property by Gale and Nikaido (1965) which crucially does not restrict the products to be Hicksian substitutes. In the second step, I use instrumental variables (IVs) to deal with endogenous prices and construct conditional moment conditions. Based on these moment conditions, the identification can be achieved under the general but high-level completeness conditions along the lines of Berry and Haile (2014).[9] To complement the discussion, I also propose low-level sufficient conditions for the identification of product-level market share functions in a mixed-logit model of demand for bundles. In particular, I show that under some common regularity conditions, product-level market share functions are identified when demand and supply shocks are jointly normal, or the data generating process (DGP) is a model of demand for multiple products across categories.

In the second part, assuming the identification of product-level market share functions, I study that of bundle-level market share functions. This is to disentangle the demand synergies among products from the unobserved correlations in the utilities of products (Gentzkow, 2007). Because only product-level market shares are observed, this task is more challenging than the usual case in which bundle-level

[7] Under linear pricing, firms set prices for single products and the price of a bundle is the sum of the prices of its single products.

[8] Under nonlinear pricing, the price of a bundle can be different from the sum of the prices of its single product, and firms can further set discounts or surcharges on the bundles of their own products.

[9] For various forms of completeness conditions, the testability, and the sufficient conditions of completeness, see Mattner (1993), D'Haultfoeuille (2011), Canay et al. (2013), Andrews (2017), Freyberger (2017), Hu and Shiu (2018).

market shares are observed. I provide both positive and negative results. First, I prove that the identification can be achieved in mixed-logit models of demand for bundle up to size two that are often used in the empirical literature.[10] Second, I show that using only product-level market shares may have limited power in identifying bundle-level market share functions in other types of models. I provide an example of non-identification in a model of demand for multiple units.

I propose a GMM estimation procedure that is similar to that in BLP models of single products. In the first step, given a guess of the demand synergy parameters and the distribution of the random coefficients, I invert the observed product-level market shares to the mean utilities of products. In the second step, I instrument out the unobserved demand shocks in the mean utilities of products and construct the GMM objective function. There are nontrivial challenges that BLP models of single products do not have. In particular, the implementation of the demand inverse is complicated due to possible Hicksian complementarities among products: when used to implement the demand inverse in the first step, the fixed-point algorithm proposed by Berry et al. (1995) may not have the contraction-mapping property and therefore may not converge. To solve this challenge, I propose to use Jacobian-based algorithms. To enhance their numerical performance, I suggest using a starting value of parameters directly constructed from the observed product-level market shares. In Monte Carlo simulations, I show that using this starting value significantly improves the numerical performance of Jacobian-based methods, reducing the convergence time by 70% compared to using standard starting value in large applications (the number of bundles being about 5000).

Finally, I illustrate the practical implementation of the methods and estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. First, the demand estimates suggest that RTE cereals and milk are overall Hicksian complementary. Moreover, the extent to which a RTE cereal product and a milk product are complementary crucially depends on the match of their characteristics (e.g. flavours, grain type, fat content), some pairs being less complementary than others. Second, I simulate a merger between a major RTE cereal producer and a milk producer. The results are aligned with Cournot (1838)'s insight: in the presence of Hicksian complementarity, mergers can be welfare enhancing. Third, I estimate the demand and replicate the same merger exercise using two alternative models: a model that completely ignores demand synergy between RTE cereals and milk, and a model that restricts all synergies to be the same across bundles of RTE cereal and milk. The first alternative model imposes that the demand of RTE cereals and that of milk are independent, predicting the merger to have no effect on welfare. The second alternative model seems to overestimate the amount of complementarity between

[10] See Gentzkow (2007), Fan (2013), Grzybowski and Verboven (2016).

RTE cereals and milk, overestimating the consumer welfare gain due to the merger.

Related Literature Empirical literature dealing with multiple purchases typically employs models of demand for bundles that rely on individual-level choice data of bundles (e.g. individual transaction data).[11] Differently, the methods in this paper only requires the availability of aggregate product-level demand data (e.g. aggregate choice probability, or sales quantities, of products) and can be applied when bundle-level demand data is not accessible. In particular, this paper is different from Iaria and Wang (2019) in three aspects. First, the methods proposed by two papers work under different data availabilities; those in this paper work when only aggregate product-level demand data is available, while those of the other paper apply when individual-level choice data of bundles is accessible. Second, identification strategies are different; in this paper, I exploit the exogenous variation in IVs to achieve identification, while that paper fully exploits the affine relationship between the utilities of the bundle and its single products due to the availability of bundle-level demand data. Third, estimation methods are different; this paper uses a GMM estimation procedure, while that paper proposes a likelihood-type estimator that resolves the dimensionality challenge due to many market-product fixed effects.

Identifying and estimating models of demand for bundles from aggregate demand at product level is a challenging task. Moreover, prices are often endogenous, which introduces additional difficulty in identification. To the best of my knowledge, this is the first paper that provides a systematic treatment of both issues in BLP-type models of demand for bundles that may exhibit Hicksian complementarity.[12] In particular, the paper proposes a novel demand inverse argument to deal with possible Hicksian complementarities among products. Berry et al. (2013) propose the connected substitutes conditions that guarantee the invertibility of the market share functions. In model of demand for bundles with only product-level market shares being available, these conditions rely on the products to be substitutes. Some papers have employed similar concepts of demand inverse of product-level market shares in model of demand for bundles. Fan (2013) studies newspaper market in

[11] Examples include consumer choice in supermarket (Hendel (1999), Dubé (2004), Lee et al. (2013), Kwak et al. (2015), Thomassen et al. (2017), Ershov et al. (2018)), household choice among motor vehicles (Manski and Sherman, 1980), choice of telecommunication services (Liu et al. (2010), Crawford and Yurukoglu (2012), Grzybowski and Verboven (2016), Crawford et al. (2018)), subscription decision (Nevo et al. (2005), Gentzkow (2007)), firms' decision on technology adoptions (Augereau et al. (2006), Kretschmer et al. (2012)). See Berry et al. (2014) for a survey of complementary choices and sections 4.2-4.3 of Dubé (2018) for a survey of econometric modelling of complementary goods.

[12] Dunker et al. (2017) also deal with price endogeneity in identification. However, instead of using the product-level market shares, they assume the availability of a vector of bundle-level market shares that has the same dimension as the number of products.

the US and assumed households subscribe to at most two newspapers. She gives sufficient conditions for the connected substitutes conditions and therefore rules out Hicksian complementarities among different newspapers. In a model of demand for multiple products across categories, Song and Chintagunta (2006) implement the demand inverse of brand-level market shares. However, they do not have theoretical results on the invertibility of the brand-level market share functions. Iaria and Wang (2019) employ the demand inverse to concentrate out fixed effects in estimation when individual-level choice data of bundles is available. In contrast, I prove the invertibility of product-level market share functions in general models of demand for bundles (e.g. mixed-logit, probit) and use the demand inverse as a key identification argument when only product-level market shares are available.

This paper also contributes to the recent literature of random-utility models of demand in the presence of multiple purchases and potentially complementarity. Fosgerau et al. (2019) employ a different approach and model Hicksian complementarity via overlapping nests. Sher and Kim (2014)'s identification arguments crucially rely on substitutes assumptions in consumers' utility,[13] while this paper does not restrict utility functions to be submodular or supermodular. Allen and Rehbeck (2019a)'s main results imply the identification of product-level market share functions in discrete choice models with additively separable heterogeneity. The following paper, Allen and Rehbeck (2019b), gives identification results of certain distributional features of the random coefficients in the case of two products (and therefore one bundle). Differently, the current paper achieves the identification using IVs and further provides identification results of bundle-level market share functions that allow for any finite number of products (and bundles). Notably, except for Fosgerau et al. (2019), all the other papers mentioned assume away endogenous prices.

Organisation In section 2, I introduce the model and necessary notations. In section 3, I motivate the model from three aspects. First, I provide various examples in the literature that can be formulated via the model. Second, I illustrate how the model can allow for Hicksian complementarity. Third, I provide Monto Carlo evidences that accent the economic relevance of the proposed model to various counterfactual analyses. In section 4, I present the main identification results. In section 5, I describe the GMM estimation procedure and its implementation. In section 6, I illustrate the practical implementation of the methods with an empirical application. Section 7 concludes. All proofs are in Appendices A-G. Figures and tables can be found in Appendix I. Additional Monte Carlo simulations can be

[13]When each consumer is assumed to consume at most one unit of each good, they impose submodularity restriction in consumers' utility (see their Assumption 2); when multi-unit demand is allowed, they use a stronger "M-natural concavity" restriction (see their Assumption 3).

found in Appendix K.

2 Model

Denote market by $t = 1, ..., T$. The definition of market depends on the concrete application. For example, one could have different geographic areas in the case of cross sectional data, or different periods in the case of panel data, or a combination of these. For individuals in market t, let $\mathbf{J}$ be the set of J market-specific products that can be purchased in isolation or in bundles. A bundle $\mathbf{b}$ is defined as a collection of single products in $\mathbf{J}$ and denote the set of available bundles in market t by $\mathbf{C}_2$.[14] Denote the outside option by 0. Individuals in market t can either choose a product $j \in \mathbf{J}$, a bundle $\mathbf{b} \in \mathbf{C}_2$, or the outside option denoted by 0. Denote by $\mathbf{C}_1 = \mathbf{J} \cup \mathbf{C}_2$ the set of available products and bundles, and by $\mathbf{C} = \mathbf{C}_1 \cup \{0\}$ the choice set of individuals in market t. Let p_{tj} denote the price of product j in market t, and $x_{tj} \in \mathbb{R}^K$ the market-product specific vector of other characteristics of j in t.

For individual i in market t, the indirect utility from purchasing product $j \in \mathbf{J}$ is:

$$\begin{aligned} U_{itj} &= u_{itj} + \varepsilon_{itj} \\ &= x_{tj}\beta_i - \alpha_i p_{tj} + \eta_{ij} + \xi_{tj} + \varepsilon_{itj} \\ &= x_{tj}^{(1)}\beta^{(1)} + x_{tj}^{(2)}\beta_i^{(2)} - \alpha_i p_{tj} + \eta_{ij} + \xi_{tj} + \varepsilon_{itj} \\ &= [x_{tj}\beta - \alpha p_{tj} + \xi_{tj}] + [x_{tj}^{(2)}\Delta\beta_i^{(2)} - \Delta\alpha_i p_{tj} + \eta_{ij}] + \varepsilon_{itj} \\ &= \delta_{tj} + \mu_{itj} + \varepsilon_{itj}, \end{aligned} \tag{1}$$

where $u_{itj} = \delta_{tj} + \mu_{itj}$ with $\delta_{tj} = x_{tj}\beta - \alpha p_{tj} + \xi_{tj}$ being market t-specific mean utility of product $j \in \mathbf{J}_t$ and $\mu_{itj} = x_{tj}^{(2)}\Delta\beta_i^{(2)} - \Delta\alpha_i p_{tj} + \eta_{ij}$ being an individual i-specific deviation from δ_{tj}, while ε_{itj} is an idiosyncratic error term. $x_{tj}^{(1)} \in \mathbb{R}^{K_1}$ is the vector of product characteristics that enter U_{itj} with deterministic coefficient(s), $\beta^{(1)}$, i.e. consumers have homogeneous taste on $x_{tj}^{(1)}$, while $x_{tj}^{(2)} \in \mathbb{R}^{K_2}$ and p_{tj} enter U_{itj} with potentially individual i-specific coefficients, $\beta_i^{(2)}$ and α_i. They capture consumers' heterogeneous tastes on characteristics $x_{tj}^{(2)}$ and sensitivities to price change. The term η_{ij} captures individual i's (unobserved) perception of the quality of product j.[15] ξ_{tj} is a market-product specific demand shock, observed to both firms and individuals but not observed to the researcher.

Throughout the paper, denote by $j \in \mathbf{b}$ the relationship of product j being in

[14] The set of products and bundles can be both market-specific, i.e. $\mathbf{J}_t$ and $\mathbf{C}_{t2}$, and the results of the paper do not change. In the empirical illustration, I will adopt market-specific $\mathbf{J}_t$ and $\mathbf{C}_{t2}$; while in the theory part, having $\mathbf{J}_t = \mathbf{J}$ and $\mathbf{C}_{t2} = \mathbf{C}_2$ greatly facilitates the exposition.

[15] Any characteristic of product j that does not vary across markets is encapsulated by the mean part of η_{ij}. Equivalently, one can specify this mean as part of $\beta^{(2)}$, i.e. product-specific intercepts in δ_{tj} and the results in this paper do not change.

bundle $\mathbf{b}$. The indirect utility for individual i in market t from purchasing products in bundle $\mathbf{b} \in \mathbf{C}_2$ is:

$$\begin{aligned} U_{it\mathbf{b}} &= \sum_{j\in\mathbf{b}} u_{itj} + \Gamma_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\ &= \sum_{j\in\mathbf{b}} (\delta_{tj} + \mu_{itj}) + \Gamma_{t\mathbf{b}} + (\Gamma_{it\mathbf{b}} - \Gamma_{t\mathbf{b}}) + \varepsilon_{it\mathbf{b}} \\ &= \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}} + \left[\sum_{j\in\mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}}\right] + \varepsilon_{it\mathbf{b}} \\ &= \delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) + \mu_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}}, \end{aligned} \tag{2}$$

where $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}) = \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$ is market t-specific mean utility of bundle $\mathbf{b}$, $\mu_{it\mathbf{b}}$ is an individual i-specific utility deviation from $\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}})$, $\Gamma_{it\mathbf{b}}$ and $\Gamma_{t\mathbf{b}}$ are the individual-market it- and market t-specific *demand synergies* among the products of bundle $\mathbf{b}$, $\zeta_{it\mathbf{b}}$ is (observed or unobserved) individual deviation from average demand synergies $\Gamma_{t\mathbf{b}}$, and $\varepsilon_{it\mathbf{b}}$ is an idiosyncratic error term. Demand synergies $\Gamma_{it\mathbf{b}}$'s capture the extra utility individual i obtains from purchasing the products in bundle $\mathbf{b}$'s in market t jointly rather than separately. Consequently, individuals may find it more (or less) appealing to purchase products jointly than separately.

Finally, the indirect utility of choosing the outside option 0 is normalized as:

$$U_{it0} = \varepsilon_{it0},$$

where ε_{it0} is an idiosyncratic error term. To complete the model, I write $\mu_{itj} = x_{tj}^{(2)}\Delta\beta_i^{(2)} - \Delta\alpha_i p_{tj} + \eta_{ij} = \mu_j(\theta_{it}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}})$ and $\mu_{it\mathbf{b}} = \sum_{j\in\mathbf{b}} \mu_{itj} + \zeta_{it\mathbf{b}} = \mu_{\mathbf{b}}(\theta_{it}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}})$ as functions of random coefficients $\theta_{it} = (\Delta\beta_i^{(2)}, \Delta\alpha_i, (\eta_{ij})_{j\in\mathbf{J}}, (\zeta_{it\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2})$ and θ_{it} is distributed according to $F \in \Theta_F$.[16] Moreover, $\varepsilon_{it} = (\varepsilon_{it0}, \{\varepsilon_{itj}\}_{j\in\mathbf{J}}, \{\varepsilon_{it\mathbf{b}}\}_{\mathbf{b}\in\mathbf{C}_2})$ are assumed to be i.i.d. according to a known continuous distribution Φ (e.g. Gumbel, Gaussian), and θ_{it} and ε_{it} are independently distributed.

Denote the vector of market t-specific mean utilities for products in $\mathbf{J}$ by $\delta_{t\mathbf{J}} = (\delta_{tj})_{j\in\mathbf{J}}$, and the vector collecting all average demand synergies by $\Gamma_t = (\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$. Define $\delta_t(\Gamma_t) = (\delta_{t\mathbf{J}}, (\delta_{t\mathbf{b}}(\Gamma_{t\mathbf{b}}))_{\mathbf{b}\in\mathbf{C}_2})$. The market share function of $\mathbf{b} \in \mathbf{C}_1$ in market

[16] Typically, the distribution of θ_{it} depends on individual i's demographic characteristics $d_i \in \mathbf{D}$. In this case, F is a mixture of distributions of $\theta_i|d_i$: $F = \sum_{d_i\in\mathbf{D}} \pi_t(d_i)F(\cdot|d_i)$, where $\pi_t(\cdot)$ is the distribution of demographics in market t.

t is:[17]

$$s_{\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F) = \int\int \mathbf{1}\{U_{it\mathbf{b}} > U_{it\mathbf{b}'} \text{ for any } \mathbf{b}' \neq \mathbf{b}, \mathbf{b}' \in \mathbf{C}\} d\Phi(\varepsilon_{it}) dF(\theta_{it})$$
$$= \int s_{\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \theta_{it}) dF(\theta_{it}), \tag{3}$$

where $s_{\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \theta_{it})$ is individual i's choice probability of $\mathbf{b}$ in market t given θ_{it}.[18] Then, product-level market share function of $j \in \mathbf{J}$ is defined as a weighted sum of the market share functions of $\mathbf{b}$'s that contain j:

$$\begin{aligned} s_{j\cdot}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma_t, F) &= \sum_{\mathbf{b} \in \mathbf{C}_1} w_{j\mathbf{b}} s_{\mathbf{b}}(\delta_t(\Gamma_t); x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F) \\ &= \mathbf{w}_j \; s(\delta_t(\Gamma_t); x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F) \end{aligned} \tag{4}$$

where $\mathbf{w}_j = (w_{j\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_1}$ is of dimension $1 \times C_1$, $w_{j\mathbf{b}}$ is the number of times j appears in $\mathbf{b}$ (and known to the researcher), and $s(\delta_t; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \theta_{it}) = (s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \theta_{it}))_{\mathbf{b} \in \mathbf{C}_1}$ is the vector of market share functions of products and bundles and of dimension $C_1 \times 1$. If $j \notin \mathbf{b}$, then $w_{j\mathbf{b}} = 0$ and the market share of bundle $\mathbf{b}$ does not contribute to the product-level market share of j; if $j \in \mathbf{b}$, then $w_{j\mathbf{b}}$ is a positive integer. When there is no bundle that contains multiple units of the same product, i.e. a situation of qualitative choice, $w_{jb} = 1$ for $j \in \mathbf{b}$. Then, (4) represents the population-level marginal choice probability of j. When a bundle contain multiple units of the same product, i.e. a situation of quantity choice, w_{jb} is equal to the units of product j purchased in the form of bundle $\mathbf{b}$ and may be greater than 1. Then, (4) represents the population-level total purchases of product j. In both cases, the aggregation in (4) is consistent with the typical product-level aggregate demand data available to the researcher.

3 Demand Synergies in Model (4)

Demand model (4) exhibits two features that are crucial in many economic analyses that models of demand for single products do not have. First, it allows for simultaneous purchases of multiple products and/or quantities. Second, it captures rich substitution patterns among products, and in particular, Hicksian complementarity. Demand synergy parameters $\Gamma_{it\mathbf{b}}$'s are the key to generate both features.

In this section, I illustrate the economic relevance of demand synergies from three aspects. First, I show that by imposing specific restrictions on $\Gamma_{it\mathbf{b}}$, model (4) covers a wide range of economic models. I provide various examples in the

[17] I abuse the expression $\mathbf{b} \in \mathbf{C}_1$ for both product j and bundle $\mathbf{b}$.

[18] Because Φ is a continuous distribution, then the event of having equal indirect utilities between two alternatives is zero.

literature and explain the economic interpretation of $\Gamma_{it\mathbf{b}}$'s in each setting. Second, I demonstrate how $\Gamma_{it\mathbf{b}}$'s generate flexible substitution patterns in demand and, in particular, Hicksian complementarity. Finally, I report Monte Carlo evidences, showing that ignoring these synergy parameters in demand estimation results in considerable bias in counterfactual analyses.

3.1 Examples and Interpretations of Demand Synergies

Example 1: Demand for Single Products, within Category This example can be seen as a particular case of (4) with $\mathbf{C}_2 = \emptyset$, or equivalently, $\Gamma_{it\mathbf{b}} = -\infty$, for all $\mathbf{b} \in \mathbf{C}_2$. This restriction on $\Gamma_{it\mathbf{b}}$ rules out simultaneous purchases of more than one product and restricts products to be substitutes.

Example 2: Demand for Multiple Products, within Category. Gentzkow (2007) considers household's choice over bundles of at most 2 different newspapers: $\mathbf{C}_2 = \{(j_1, j_2) : j_1 < j_2, j_1, j_2 \in \mathbf{J}\}$ and $\mathbf{C} = \{0\} \cup \mathbf{J} \cup \mathbf{C}_2$. In general, one can allow for choice over bundles of up to L different products: $\mathbf{C}_2 = \{(j_1, ..., j_l) : j_1 < ... < j_l, j_1, ..., j_L \in \mathbf{J}\}$. As shown in Iaria and Wang (2019), demand synergy $\Gamma_{it\mathbf{b}}$ can proxy, for example, preferences for variety, synergies in consumption.

Example 3: Demand for Multiple Products, across Categories. Grzybowski and Verboven (2016) and Ershov et al. (2018) consider purchases of products across different categories. In the simplest case in which a bundle is defined as a collection of 2 different products (chips and soda) with each belonging to a different category (salty snacks and carbonated drinks), we have $\mathbf{C}_2 = \mathbf{J}_1 \times \mathbf{J}_2 = \{(j_1, j_2) : j_1 \in \mathbf{J}_1, j_2 \in \mathbf{J}_2\}$. In the example of potato chips and carbonated soda (Ershov et al., 2018), $\Gamma_{it\mathbf{b}}$'s are interpreted as synergies in consumption.

Example 4: Quantity Choice, Multiple Units. As a deviation from Example 1, individuals purchase not only one out of J products but also a discrete quantity of the chosen product. This can be captured by $\mathbf{C}_2 = \{(j, ..., j) : j \in \mathbf{J}, \text{the length of } (j, ..., j) \leq L.\}$, where L is the maximal units individuals can purchase. In the simplest case, individuals can purchase the outside option 0, a unit of product $j \in \mathbf{J}$ (single product), or a bundle of two identical units (j, j), $j \in \mathbf{J}$. Demand synergy $\Gamma_{it(j,j)}$ is then interpreted as extra utility individual i obtains from purchasing an additional unit of product j relative to the first unit: $\Gamma_{it(j,j)} < 0 (> 0)$ implies a decreasing (increasing) marginal utility of purchasing product j. If $\Gamma_{it(j,j)} = 0$, then individual i's utility from purchasing the second unit of product j remains the same as that from purchasing the first unit.

Example 5: Multiple Discreteness. Demand model of multiple discreteness (see Hendel (1999) and Dubé (2004)) can be seen as an extension of Example 4 that further includes bundles defined as a collection of multiple units of different products: $\mathbf{b} = (\ \underbrace{(j, ..., j)}_{n_j}\ )_{j \in \mathbf{J}}$, where n_j is the number of units of product j. As shown in Iaria and Wang (2019) (Appendix 8.1), Dubé (2004)'s model of multiple discreteness can be formulated by specifying $\Gamma_{it\mathbf{b}} = \sum_{j \in \mathbf{J}} \Gamma_{it(j,...,j)}$, where $\Gamma_{it\underbrace{(j, ..., j)}_{n_j}} \leq 0$ for any $n_j > 1$ and $j \in \mathbf{J}$. The non-positive $\Gamma_{it(j,...,j)}$ represents non-increasing marginal utility of consuming additional units of product j during one consumption moment and the additivity in $\Gamma_{it\mathbf{b}}$ across $j \in \mathbf{J}$ represents the independence between consumption moments.[19]

Example 6: Multi-Category Multi-Store Demand. Thomassen et al. (2017) study a multi-category multi-store demand model, in which individual purchases multiple units in each of K product categories and purchase all the units of the same category in the same store. Consider a simple case in which individual purchases at most one unit in each of 2 product categories (k_1 and k_2) from 2 stores (S_1 and S_2). This can be captured by $\mathbf{J} = \{j = (j^1, j^2) : j^1 = k_1, k_2,\ j^2 = S_1, S_2\}$ and $\mathbf{C}_2 = \{(j, r) : j, r \in \mathbf{J},\ j^1 \neq r^1\}$. A product is defined as a Cartesian product of categories and stores with first coordinate being category and the second being store (category 1 in store 2) and a bundle is defined as a Cartesian product of two products that differ in their first coordinate (category 1 in store 2 and category 2 in store 2). Demand synergy $\Gamma_{it(j,r)}$ is interpreted as shopping cost due to store choice: $\Gamma_{it(j,r)} = 0$ if $j^2 = r^2$ (one-stop shopping, buy products of both categories from the same store), and negative otherwise (multi-stop shopping, purchase products in one category from store 1 and those in the other category from store 2).

3.2 Hicksian Substitutions and Demand Synergies

Using a mixed-logit model of demand for multiple products across categories (see Example 3 of section 3), I illustrate how the signs of cross-price elasticities in model (4), i.e. Hicksian substitutability (positive cross-price elasticities) or complemen-

[19] Due to Dubé (2004)'s perfect substitute specification, individual will consume up to one product during one consumption moment.

tarity (negative cross-price elasticities) are determined by demand synergies $\Gamma_{it\mathbf{b}}$.[20]

To ease exposition, I drop the notation of market t and product characteristics in product-level market share functions. Let's consider the cross-price elasticity between products $j \in \mathbf{J}_1$ and $r \in \mathbf{J}_2$:[21]

$$\epsilon_{jr} = \frac{p_r}{s_{j.}} \int \alpha_i [s_{j.}(\delta(\Gamma); \theta_i) s_{r.}(\delta(\Gamma); \theta_i) - s_{jr}(\delta(\Gamma); \theta_i)] dF(\theta_i),$$

where $\alpha_i > 0$. Different from models of demand for single products, the cross-price elasticity ϵ_{jr} has an additional term $-s_{jr}(\delta(\Gamma); \theta_i)$. When this term is relatively large, i.e. the joint purchase probability for products j and r is relatively large, we may have a negative ϵ_{jr}, i.e. Hicksian complementarity between j and r. In the case of two products and one bundle, i.e. $\mathbf{J} = \{1, 2\}$ and $\mathbf{C}_2 = \{(1, 2)\}$, Gentzkow (2007) shows that $\Gamma_{jr} = 0$ is the cut-off value for Hicksian substitutability and complementarity: $\varepsilon_{12} < 0$ if and only if $\Gamma_{(1,2)} > 0$. When there are more than 2 products, even though $\Gamma_{(j,r)} = 0$ may not be the cut-off value for Hicksian substitutability and complementarity between j and r, the intuition remains similar. Note that whether j and r are substitute, complementary or independent, i.e. $\epsilon_{jr} > 0$, $\epsilon_{jr} < 0$ or $\epsilon_{jr} = 0$, is determined by the weighted average of $s_{ij.}s_{ir.} - s_{ijr}$. The latter is further determined by the magnitude of synergy between j and r, Γ_{jr}, relative to other demand synergies. If Γ_{jr} is sufficiently negative, then s_{ijr} is close to zero and thus $\epsilon_{jr} > 0$. As an extreme case, when $\Gamma_{jr} = -\infty$, i.e. bundle (j, r) is not in the choice set, j and r are always substitute and therefore ϵ_{jr} is positive. If Γ_{jr} is positive and large enough relative to $\Gamma_{j'r'}$ for all $(j', r') \neq (j, r)$, then $s_{ij.} - s_{ijr}$ and $s_{ir.} - s_{ijr}$ are negligible relative to s_{ijr}. Then, the sign of ϵ_{jr} is determined by the population average of $s_{ijr}^2 - s_{ijr}$. Since s_{ijr} is strictly between 0 and 1, $s_{ijr}^2 - s_{ijr}$ is always negative and therefore $\epsilon_{jr} < 0$. If Γ_{jr} takes some medium value in $(-\infty, \infty)$, we can expect $\epsilon_{jr} = 0$ and j and r are independent.

3.3 Ignoring Demand Synergies in Demand Estimation: Bias in Counterfactual Simulations

I provide Monte Carlo evidences and show that ignoring demand synergies in demand estimation potentially leads to substantial bias in counterfactual analyses.

[20] In both (1) and (2), income effect is ruled out (or enters linearly in the direct utilities of all alternatives). Consequently, in most part of the paper, Hicksian complementarity (substitutability) coincides with negative (positive) cross-price elasticities. One can adopt income effect by using a different specification of indirect utilities. For example, one specification in Berry et al. (1995) is that price p_j enters the indirect utility via $-\alpha_i \log(y_i - p_j)$, where y_i is individual i's income. Note that if $p_j \ll y_i$, i.e. the income is much larger than product price, then $-\alpha_i \ln(y_i - p_j) \approx -\alpha_i \ln y_i - \frac{\alpha_i}{y_i} p_j$, and the specification in (1) and (2) is suitable.

[21] See Appendix A for details of the computation.

To simplify the exposition while still capturing the economic essence, in each of the scenario below, I will suppose that the data generating process (DGP) is a multinomial-logit model of demand for bundles.[22] Consider a simple setting in which there are two product categories $\mathbf{J}_1 = \{1,2\}$ and $\mathbf{J}_2 = \{3,4\}$. An individual can purchase a product $j \in \mathbf{J}_1 \cup \mathbf{J}_2$, or a bundle (j_1, j_2), where $j_1 \in \mathbf{J}_1$ and $j_2 \in \mathbf{J}_2$, or the outside option denoted by 0. Each product $j \in \mathbf{J}_1 \cup \mathbf{J}_2$ has a price p_j and a characteristic x_j, and the price of a bundle (j_1, j_2) is equal to $p_{j_1} + p_{j_2}$, i.e. there is neither discount nor surcharge. For each bundle (j_1, j_2), the corresponding demand synergy is $\Gamma_{j_1 j_2}$, constant across individuals. Demand synergy parameters are not zero in the true DGP of each scenario below. In addition, I estimate a model of demand that ignores synergies in demand and imposes $\Gamma_{j_1 j_2} = 0$ for all $j_1 \in \mathbf{J_1}$ and $j_2 \in \mathbf{J}_2$. Note that this estimated model is equivalent to two independent models of demand for single products in $\mathbf{J}_1$ and in $\mathbf{J}_2$. I use this estimated model to predict the counterfactuals in each scenario and compare the outcomes to those predicted by the true model. For details of the DGPs, see Appendix J.

Multi-category demand. First, I consider multi-category demand in which products in $\mathbf{J}_1$ and $\mathbf{J}_2$ are complementary in consumption (i.e. demand synergies are positive). Suppose that all synergy parameters are equal to Γ. Moreover, there are an equal number of producers to that of products, each producing one product. I simulate two counterfactual scenarios: cross-category merger and tax on products in $\mathbf{J}_1$. Table 1 summarise the results.

In the upper panel, I simulate a merger between the producer of product $1 \in \mathbf{J}_1$ and the producer of $3 \in \mathbf{J}_2$. The outcomes predicted by the true model (columns 1 and 3) confirm Cournot (1838)'s intuition: the merged producer internalises complementarity between 1 and 3 and prices decrease after the merger, enhancing consumer surplus. Moreover, the more products are complementary, the more consumer surplus is enhanced after the merger. In contrast, the estimated model imposes $\Gamma = 0$ and restricts the demand of products in $\mathbf{J}_1$ and $\mathbf{J}_2$ to be independent, ruling out the complementarity. As a result, the estimated model predicts neither an increase nor a decrease in prices after the merger (columns 2 and 4).

In the lower panel, I simulate a scenario in which the government imposes a 25-cent tax on the prices of products in $\mathbf{J}_1$. Excise tax may have both positive and negative consequences on consumers' welfare. As pointed out by Dubois et al.

[22] There are two reasons for this choice. First, a multinomial-logit model of demand for bundles rules out unobserved correlation in utilities of products and therefore accents the economic consequences of synergies in demand. Second, I will estimate a model of demand that rules out the demand synergies using IV's, and compare the counterfactual outcomes predicted by this estimated model to those by the true model. Using a multinomial-logit DGP makes the estimation procedure transparent (a linear IV regression), and minimises econometric errors.

(ming), on one hand, soda taxes may considerably reduce sugar consumption. Consumers then benefit from the averted internalities achieved by the tax (e.g. less health problem in the future); on the other hand, there is also a direct consumer welfare loss from higher prices induced by the tax. In the recent literature using discrete-choice models to study the impact of excise tax, this direct welfare loss is quantified by models of demand for single products, and in most cases, within a product category.[23] However, these models may not take into account the externalities of taxes on the consumption of products in other categories. Depending on the shape of multiple-category demand, this could under or over-estimate the welfare loss, biasing the evaluation of net effect of taxes on consumer welfare.

As shown in the last row of Table 1, in the presence of synergy in consumption, the direct consumer welfare loss due to the tax is substantially underestimated by the estimated model. When the synergy in consumption (Γ) is moderate (columns 1 and 2), the estimated model underestimates the consumer welfare loss by about 13.5%;[24] when Γ is large (columns 3 and 4), the estimated model underestimates the consumer welfare loss by about 25.5%. Intuitively, the estimated model with $\Gamma = 0$ switches off externality of the tax on the consumption of products in $\mathbf{J}_2$. Even if the predicted price and demand changes in category $\mathbf{J}_1$ are quantitatively similar across the true and the estimated models, the estimated model fails to predict the decreasing demand in category $\mathbf{J}_2$, underestimating the consumer surplus loss.

Table 1: Counterfactual simulations, multi-category demand

Synergy in consumption	Moderate ($\Gamma = 2$)		Large ($\Gamma = 5$)	
Model	True	Estimated	True	Estimated
	$\Gamma = 2$	$\Gamma = 0$	$\Gamma = 5$	$\Gamma = 0$
Merger across categories				
Δp, cat.1	−1.27%	0%	−1.55%	0%
Δp, cat.2	−2.19%	0%	−1.11%	0%
Δ Consumer Surplus	5.88%	0%	6.24%	0%
25-cent tax on prods. in cat. 1				
Δp, cat.1	11.61%	11.55%	8.80%	8.59%
Δs, cat.1	−23.54%	−24.20%	−16.31%	−16.77%
Pass-through	87.10%	86.77%	87.84%	86.38%
Δp, cat.2	−0.72%	0%	−1.04%	0%
Δs, cat.2	−7.10%	0%	−6.41%	0%
Δ Consumer Surplus	−16.08%	−13.91%	−14.85%	−11.07%

[23]See Bonnet and Réquillart (2013), Wang (2015), Griffith et al. (2018, 2019), Dubois et al. (ming) among others. Note that Dubois et al. (ming) employs a model of demand for non-alcoholic drinks with an outside option that include snacks.

[24]The number is obtained using the welfare losses in columns 1 and 2: 13.5% = (16.08% − 13.91%)/16.08%.

Supermarket competition. Second, I consider supermarket competitions. Suppose that there are two supermarkets, S_1 and S_2. S_1 owns product $1 \in \mathbf{J}_1$ and $3 \in \mathbf{J}_2$; S_2 owns product $2 \in \mathbf{J}_1$ and $4 \in \mathbf{J}_2$. In the factual scenario, S_1 and S_2 simultaneously set prices for their own products and implement linear pricing. I simulate two counterfactual scenarios: a merger between S_1 and S_2, and duopoly under nonlinear pricing. Table 2 summarises the results.

In the upper panels, I suppose that S_1 and S_2 are geographically distant and buying products from different supermarkets incurs a shopping cost: $\Gamma_{14} = \Gamma_{23} = \Gamma < 0$; purchasing products from the same supermarket does not incur such shopping cost: $\Gamma_{13} = \Gamma_{24} = 0$. As argued in Thomassen et al. (2017), individuals may prefer to purchase products of different categories from one shop, rather than from different shops (captured by Γ). This behaviour of one-stop shopping can generate complementary cross-category pricing effects and may have different implications for supermarket competition from the multiple-stop shopping behaviour. Consequently, the extent to which individuals are one-stop shoppers, i.e. the magnitude of Γ, is crucial when studying supermarket competition; using a model of demand that ignores shopping cost Γ may lead to substantial bias. I simulate a merger between S_1 and S_2, and find that the estimated model imposing $\Gamma = 0$ (i.e. ignoring the shopping cost) substantially underestimates price increases in both categories, and therefore underestimates consumer welfare loss due to the merger.[25] When the shopping cost is large (columns 3 and 4), the consumer welfare loss is almost underestimated by 50%. As found by Thomassen et al. (2017), in the presence of shopping cost ($\Gamma < 0$) and supermarket competition, one-stop shoppers may create cross-category complementarity in pricing and have greater pro-competitive impact. Once S_1 and S_2 are merged, purchasing products from the same or different supermarkets is irrelevant to the profit of the merged firm and this pro-competitive effect disappears. The estimated model imposes $\Gamma = 0$, ruling out this pro-competitive effect from one-stop shoppers. As a result, the consumer surplus loss due to the merger is underestimated.

In the lower panels, I suppose that S_1 and S_2 are geographically close (e.g. express stores in city centre) and study nonlinear pricing competition when consumers purchase products from complementary categories ($\Gamma_{j_1 j_2} = \Gamma > 0$, e.g. sandwich and soft drink). It is largely acknowledged that supermarkets offer bundles of (complementary) products with a discount and there may exist several rationales for the bundling behaviour, with or without complementarity in demand.[26] The welfare implication of nonlinear pricing competition is ambiguous and crucially depends

[25] Both supermarkets are still physically distant after the merger and the synergy parameters remain the same.

[26] See Adams and Yellen (1976), McAfee et al. (1989), Matutes and Regibeau (1992), Armstrong and Vickers (2010), Armstrong (2016a) among others.

on the shape of demand (Armstrong and Vickers, 2010). As shown in the bottom panel, in both scenarios of moderate (column 1) and large synergies (column 3) in consumption, the true model predicts an increase in prices of single products (consumer surplus loss when consumers purchase single product), but a discount on the bundles provided by the same supermarket (consumer surplus gain when consumers purchase bundles of products from the same supermarket). The net welfare effect is then determined by the magnitude of synergy in consumption, i.e. Γ: when Γ is large, consumers tend to purchase more bundles from the same supermarket (relative to single product) and therefore there is an overall consumer welfare gain (column 3); when Γ is small or moderate, more consumers tend to purchase single product and therefore there is an overall consumer welfare loss (column 1). The estimated model (columns 2 and 4) imposes $\Gamma = 0$ and predicts too much purchase of single products, amplifying the consumer welfare loss. This leads to overestimate consumer surplus loss when Γ is moderate, and underestimate consumer surplus gain when Γ is large. Moreover, the profit change predicted by the estimated model is also misleading (last row); even the sign can be wrongly predicted (column 2).

Table 2: Counterfactual simulations, supermarket competition

Shopping cost	Moderate ($\Gamma = -2$)		Large ($\Gamma = -5$)	
Model	True	Estimated	True	Estimated
	$\Gamma = -2$	$\Gamma = 0$	$\Gamma = -5$	$\Gamma = 0$
Supermarket merger				
Δp, cat.1	15.76%	8.61%	17.49%	8.56%
Δp, cat.2	15.95%	8.77%	17.69%	8.72%
Δ Consumer Surplus	−22.36%	−12.35%	−24.01%	−12.27%
Synergy in consumption	Moderate ($\Gamma = 2$)		Large ($\Gamma = 5$)	
Model	True	Estimated	True	Estimated
	$\Gamma = 2$	$\Gamma = 0$	$\Gamma = 5$	$\Gamma = 0$
Nonlinear pricing				
Δp, cat.1	11.10%	9.64%	10.70%	12.08%
Δp, cat.2	11.12%	9.66%	10.61%	12.00%
Discount	16.34%	17.65%	14.02%	16.54%
Δ Consumer Surplus	−0.29%	−0.46%	6.34%	2.99%
Δ Profit	−0.79%	2.25%	−8.38%	−5.62%

4 Identification

I first give the assumptions the identification and estimation will rely on.

Assumption 1. *For any* $t \in \mathbf{T}$*,* *(i). (Data availability) Product-level market shares, $\mathcal{s}_{tj.} = \sum_{\mathbf{b}\in\mathbf{C}_1} w_{j\mathbf{b}}\mathcal{s}_{t\mathbf{b}}$, are observed to the researcher for $j \in \mathbf{J}$.*

(ii). (Mix and match) If bundle $\mathbf{b} \in \mathbf{C}_2$, then $j \in \mathbf{J}$, for any $j \in \mathbf{b}$.

(iii). (Many-market) The total number of products, $|\mathbf{J}|$, and bundles $|\mathbf{C}_2|$, are fixed while the number of markets, $T = |\mathbf{T}|$, is large.

Assumption 1(i) specifies the data environment in which only product-level (rather than bundle-level) market shares are available to the researcher. To simplify the exposition, I assume that product-level market shares are observed without error, i.e. the number of individuals in each market is sufficiently large. In estimation, one can allow for the number of individuals to increase fast enough relative to the number of markets; the results of the paper still hold.[27] Assumption 1(ii) clarifies that bundle is a result of individuals' multiple purchases, i.e. a bundle is defined as a set of products purchased by individuals. The definition of product may vary from application to application. If some single products are always sold together (e.g. business-class flight is only available via bundle of business-class seat and large allowance of luggage), as long as purchase data of such combinations is available, i.e. Assumption 1(i) holds, then one can define such combination as a product and Assumption 1(ii) is not violated. Assumption 1(iii) focuses on the *many-market* setting in which the numbers of products and bundles are fixed while the number of markets increases asymptotically.

As a result of Assumption 1(iii), without further restrictions, the number of demand synergy parameters to be identified (i.e. Γ_t for all $t \in \mathbf{T}$) increases as T increases. This challenge of dimensionality introduces substantial difficulty in identification and estimation.[28] To overcome this challenge, I propose the following assumption along the lines of Gentzkow (2007)'s model of demand for bundles (and also its generalised version in Iaria and Wang (2019)):

Assumption 2. *For any $\mathbf{b} \in \mathbf{C}_2$ and $t \in \mathbf{T}$,*

$$\Gamma_{t\mathbf{b}} = g(x_{t\mathbf{b}}; \Sigma_g) + \Gamma_{\mathbf{b}},$$

where $x_{t\mathbf{b}}$ a vector of observed market-bundle specific non-price characteristics, $g(\cdot; \Sigma_g)$ a function of $x_{t\mathbf{b}}$ parametrized by and continuously differentiable with respect to $\Sigma_g \in \Theta_{\Sigma_g}$, and $\Gamma_{\mathbf{b}}$ is a bundle-specific fixed effect.

[27] In models of demand for single products, Freyberger (2015) allows for sampling errors in the observed market shares. He shows the consistency and asymptotic normality of the GMM estimator by requiring the number of individuals to increase fast enough relative to the number of markets.

[28] In particular, these parameters become incidental in estimation. Providing a solution to this problem is beyond the scope of the current paper.

Assumption 2 reduces the dimension of demand synergy parameters to the sum of $\dim(\Sigma_g)$ and $\dim(\Gamma) = \dim((\Gamma_{\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}) = |\mathbf{C}_2|$, which remains *fixed* as T increases. The main motivation for this assumption is that bundle-level market shares are not observed to the researcher. If they were all available, then one could directly identify and estimate model (3), rather than model (4), à la BLP with bundle-level instruments. In this case, Assumption 2 is redundant.

Different from the model used in Iaria and Wang (2019), Assumption 2 assumes linear pricing in the factual, i.e. the observed price of a bundle is the sum of the prices of its single products. This excludes market-specific nonlinear pricing in the factual, i.e. bundle-specific discounts or surcharges. While it is possible to extend the main results of the paper to allow for nonlinear pricing in the factual, I focus on the situations covered by Assumption 2 and will explore this extension in future research. Note that Assumption 2 does not rule out the possibility of simulating counterfactuals under nonlinear pricing. In such scenarios, this assumption implies that the source of unobserved variations across markets is limited to the market-product specific demand shocks $\xi_{t\mathbf{J}}$. As a result, conditional on the observed characteristics of products and bundles, prices vary across markets only due to variations in $\xi_{t\mathbf{J}}$.

Assumption 2 summarises situations with or without exogenous characteristics of bundles. Both situations can be similarly treated in the discussion of identification and estimation. To simplify the exposition, I will focus on the leading case $g \equiv 0$, i.e. $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}}$.

4.1 Economic Analyses and Sufficient Demand Primitives

Under Assumptions 1 and 2, the identification problem the researcher faces is the following: for any $j \in \mathbf{J}$ and $t \in \mathbf{T}$,

$$\mathcal{s}_{tj\cdot} = s_{j\cdot}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F), \tag{5}$$

where $\delta_{t\mathbf{J}} = (\delta_{tj})_{j\in\mathbf{J}}$ with $\delta_{tj} = x_{tj}\beta - \alpha p_{tj} + \xi_{tj}$, $(\mathcal{s}_{tj\cdot}, x_{tj}, p_{tj})$ is observed to the researcher, and $(\alpha, \beta, \Gamma, F)$ are the structural parameters. Identifying $(\alpha, \beta, \Gamma, F)$ does enable to conduct most economic analyses in practice. However, this may be overly sufficient and challenging.

In this section, I characterise two sets of demand primitives that are respectively sufficient for two classic types of economic analyses in model of demand for bundles: those under linear pricing and under nonlinear pricing. Under linear pricing strategy, firms set prices of their single products; the price of a bundle is defined as the sum of the prices of its single products. Under nonlinear pricing strategy, firms can not only set prices of its single products but also the bundles of their own products. Then, the price of a bundle can be different from the sum of the prices of its products when there is a discount or surcharge. The proposed sets of demand primitives are "coarser" than $(\alpha, \beta, \Gamma, F)$ and therefore easier to be identified. The following proposition sheds light on the set of demand primitives needed to conduct each of the two types of analyses in merger simulations.[29]

Proposition 1. *Suppose that Assumptions 1-2 and condition 4 hold, $\alpha_i > 0$, and the observed prices $p_{t\mathbf{J}}$ (and also those after mergers) are uniquely generated from a simultaneous Bertrand price-setting game under complete information with constant marginal cost c_{tj} for $j \in \mathbf{J}$.*

- *Suppose that (α, β) are identified and $s_{t\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F)$ are identified as functions of $(\delta_{t\mathbf{J}}, x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}})$. Then,*
  - *price elasticities at $p_{t\mathbf{J}}$ are identified.*
  - *c_{tj}'s are identified.*
  - *assuming linear pricing after the merger, the changes of prices, profits, consumer surplus, social welfare are identified.*
- *Suppose that (α, β, Γ) are identified and $c_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} c_{tj}$ for $\mathbf{b} \in \mathbf{C}_{t2}$. If one of the following conditions holds:*
  - *$\alpha_i = \alpha$, and $s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F)$ is identified as a function of $(\delta_t, x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}})$, for any $\mathbf{b} \in \mathbf{C}_1$.*
  - *F is identified,*

  then, assuming nonlinear pricing after the merger, the changes of prices, profits, consumer surplus, social welfare are identified.

Proof. See Appendix B. □

Remark 1. *The condition $c_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} c_{tj}$ implies that there is no additional cost for firms to set bundle-specific prices. The second statement of Proposition 1 still holds if there is such additional cost and it is known to the researcher.*

The take-away of Proposition 1 is clear: identifying (α, β) and $s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F)$ (as functions of $(\delta_{t\mathbf{J}}, x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}})$) is already enough for merger simulations under linear pricing; for those under nonlinear pricing, one has to further separately identify Γ and F (or $s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F)$, as a function of $(\delta_t, x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}})$ for all $\mathbf{b} \in \mathbf{C}_2$). In the next section, I will discuss the identification of model (4) in two parts. In the first part, I discuss the identification of (α, β) and $s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F)$; in the second part, assuming the identification of (α, β) and $s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma, F)$, I study the separable identification of Γ and F (or $s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F)$ for $\mathbf{b} \in \mathbf{C}_2$).

[29]Proposition 1 also applies to analyses other than merger simulations.

4.2 Identification of Product-Level Market Share Functions

I follow the classical approach in demand models of aggregate market shares and use a two-step identification strategy. In the first step, I recover the mean utilities of products using a novel demand inverse to deal with possible Hicksian complementarities among products; in the second step, I construct moment conditions using IVs to deal with endogenous prices and identify product-level demand primitives.

Demand Inverse in Model (4) **with Complementarity.** The first step hinges on the invertibility of product-level market share functions:

Theorem 1. *Suppose that Assumption 1-2 holds. Then, under regularity condition 4 in Appendix C, for any* (Γ', F')*, there exists at most one* $\delta'_{t\mathbf{J}}$ *such that* $s_{\mathbf{J}\cdot}(\delta'_{t\mathbf{J}}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma', F') = \mathscr{s}_{t\mathbf{J}\cdot}$.

Proof. See Appendix C. □

When (Γ', F') are the true parameters (Γ, F), the vector of the true mean utilities of products, $\delta_{t\mathbf{J}}$, is the unique solution of $s_{\mathbf{J}\cdot}(\delta'_{t\mathbf{J}}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F) = \mathscr{s}_{t\mathbf{J}\cdot}$. As a result, the function $s_{\mathbf{J}\cdot}(\cdot; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F)$ is globally invertible. Denote its inverse by:

$$\delta_{t\mathbf{J}} = s^{-1}_{\mathbf{J}\cdot}(\mathscr{s}_{t\mathbf{J}\cdot}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F). \tag{6}$$

Iaria and Wang (2019) prove a similar result in a mixed logit model of demand for bundles and use it to reduce the dimensionality of fixed effects in a likelihood estimation procedure. Theorem 1 differs from theirs in two aspects. First, Theorem 1 applies to any random utility model that satisfies Assumptions 1-2 and condition 4. The mixed-logit model of demand is a particular case. Second, this invertibility result is used as a fundamental identification argument in this paper, and is the key step to construct moment conditions when only product-level market shares are available.

Importantly, the demand inverse in Theorem 1 is different from the classical demand inverse in demand models of single products in two ways. First, the invertibility in Theorem 1 builds on different arguments. In general, the invertibility of market share functions in demand models of single products follows from the connected substitutes conditions (Berry et al., 2013). These conditions may not apply to model (4) because products can be Hicksian complementary.[30] Instead, the invertibility of product-level market share functions in Theorem 1 builds on the affine relationship between the utilities of bundles and single products (see equation (2)) and on the

[30]Moreover, it seems hard to find a transformation of product-level market share functions in model (4) under which the transformed market shares functions satisfy the connected substitutes conditions.

P-matrix property by Gale and Nikaido (1965), which-crucially-does not require the products to be Hicksian substitutes. Second, the demand inverse in Theorem 1 may not be implemented by the fixed-point contraction mapping algorithm proposed by Berry et al. (1995). This is because the contraction mapping property of the algorithm may not hold when (some) products are Hicksian complementary in model (4). I propose to use Jacobian-based algorithms to implement this demand inverse.[31] See section 5.2 for details of the implementation.

When $(\Gamma', F') \neq (\Gamma, F)$, it is possible that there is no $\delta'_{t\mathbf{J}}$ such that $s_{\mathbf{J}.}(\delta'_{t\mathbf{J}}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma', F') = \mathscr{s}_{t\mathbf{J}.}$.[32] In this case, such (Γ', F') are directly ruled out of the identification set of (Γ, F). The following identification discussion will restrict to (Γ', F') such that $\delta'_{t\mathbf{J}}$ exists.

Instrumental Variable Approach. Combining equation (6) and $\delta_{t\mathbf{J}} = x_{t\mathbf{J}}\beta - \alpha p_{t\mathbf{J}} + \xi_{t\mathbf{J}}$, I obtain:

$$x_{t\mathbf{J}}\beta - \alpha p_{t\mathbf{J}} + \xi_{t\mathbf{J}} = s^{-1}_{\mathbf{J}.}(\mathscr{s}_{t\mathbf{J}.}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F). \tag{7}$$

The source of price endogeneity is $\xi_{t\mathbf{J}}$: $\xi_{t\mathbf{J}}$ are observed to firms and therefore $p_{t\mathbf{J}}$ are set based on $\xi_{t\mathbf{J}}$. Consequently, $p_{t\mathbf{J}}$ and $\xi_{t\mathbf{J}}$ are correlated, while $\xi_{t\mathbf{J}}$ are not observed to the researcher. Beyond the price endogeneity, Γ and F constitute parameters that cannot be pinned down without further assumption. I use IVs to solve these challenges:

Assumption 3. *There are random variables $z_{t\mathbf{J}} = (z_{tj})_{j \in \mathbf{J}}$, such that $\mathbb{E}[\xi_{t\mathbf{J}} | z_{t\mathbf{J}}, x_{t\mathbf{J}}] = 0$ almost everywhere.*

Assumption 3 gives rise to conditional moment restrictions:

$$\mathbb{E}\left[\xi_j(\beta, \alpha, \Gamma, F; \mathscr{s}_{t\mathbf{J}.}, x_{t\mathbf{J}}, p_{t\mathbf{J}}) | z_{t\mathbf{J}}, x_{t\mathbf{J}}\right] = 0 \quad a.e., \tag{8}$$

for $j \in \mathbf{J}$, where $\xi_j(\beta, \alpha, \eta, \Gamma, F; \mathscr{s}_{t\mathbf{J}.}, x_{t\mathbf{J}}, p_{t\mathbf{J}}) = s^{-1}_j(\mathscr{s}_{t\mathbf{J}.}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F) - x_{tj}\beta + \alpha p_{tj}$. The identification of $s_{\mathbf{J}.}(\delta_{t\mathbf{J}}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F)$ (or equivalently its inverse $s^{-1}_{\mathbf{J}.}(\mathscr{s}_{t\mathbf{J}.}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma, F)$) by moment conditions (8) can follow from general arguments in nonlinear models using IVs. In demand models of single products, one can leverage completeness conditions of the joint distribution of $(z_{t\mathbf{J}}, x_{t\mathbf{J}}, \mathscr{s}_{t\mathbf{J}.}, p_{t\mathbf{J}})$ with respect to $(\mathscr{s}_{t\mathbf{J}.}, p_{t\mathbf{J}})$ (Berry and Haile, 2014). Intuitively, this requires sufficiently rich variation in $(z_{t\mathbf{J}}, x_{t\mathbf{J}})$

[31] Conlon and Gortmaker (2020) provide a review of numerical methods for implementation of demand inverse in demand models of single products.

[32] For example, if the DGP is such that the sum of observed product-level market shares is larger than one, then any demand models of single products ($\Gamma_{\mathbf{b}} = -\infty$ for any $\mathbf{b}$) cannot rationalise the observed product-level market shares and hence the demand inverse is not feasible with $\Gamma' = -\infty$.

that can distinguish any function of the endogenous variables $(s_{t\mathbf{J}\cdot}, p_{t\mathbf{J}})$ from others. In the context of (8), the same general arguments also apply. One needs variation in $(z_{t\mathbf{J}}, x_{t\mathbf{J}})$ to distinguish $\xi_{\mathbf{J}}(\beta, \alpha, \Gamma, F; \cdots)$ from $\xi_{\mathbf{J}}(\beta', \alpha', \Gamma', F'; \cdots)$ for any $(\beta', \alpha', \Gamma', F') \neq (\beta, \alpha, \Gamma, F)$. As long as such variation is available, having demand synergy parameters Γ does not conceptually introduce additional difficulty for identification.

Despite the generality, these arguments and required conditions are often high-level. In what follows, I propose *low-level* sufficient conditions for the identification of product-level market share functions in the case of mixed-logit models of demand for bundles. In the following, I will focus on cost-type variables.[33] In Appendix L, I propose similar sufficient conditions for other commonly used instruments (e.g. BLP-type instruments, exogenous product characteristics).

Denote by $\mathbf{D}_x^{(1)}$ the support of $x_{t\mathbf{J}}^{(1)}$ and by $\mathbf{D}_x^{(2)}$ the support of $x_{t\mathbf{J}}^{(2)}$. Moreover, both $\mathbf{D}_x^{(1)}$ and $\mathbf{D}_x^{(2)}$ open. Suppose that the ownership of each product is the same across markets and that prices are generated from a simultaneous Bertrand pricing game under complete information with constant marginal cost c_{tj}, for $j \in \mathbf{J}$. Without loss of generality, I specify $c_{tj} = z_{tj} + w_{tj}$, where z_{tj} is cost shifter for product j and w_{tj} is exogenous supply shock that is observed to firms but not observed to the researcher. The main identification result of the product-level market share functions is the following:

Theorem 2. *Suppose that Assumptions 1-3 and regularity condition 1 in Appendix D holds. Moreover, the following conditions hold:*

1. $(x_{t\mathbf{J}}, z_{t\mathbf{J}})$ *is independent of* $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$*, the support of* $z_{t\mathbf{J}}$ *is* $\mathbb{R}^J$.
2. $\alpha_i = \alpha > 0$
3. *Given* $x_{t\mathbf{J}}^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x_{t\mathbf{J}} + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)})$ *is a continuous function of* $(\beta x_{t\mathbf{J}} + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$.
4. F *has compact support.*

Then,

- *If* $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ *is Gaussian distributed, then* (α, β) *is identified and* $s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x_{\mathbf{J}}^{(2)}, p_{\mathbf{J}}, \Gamma, F)$ *are identified as functions of* $(\delta_{\mathbf{J}}, x_{\mathbf{J}}^{(2)}, p_{\mathbf{J}}) \in \mathbb{R}^J \times \mathbf{D}_x^{(2)} \times \mathbb{R}^J$.
- *If the DGP is a model of demand for multiple products across categories (see section 3.1), then under regularity condition 2 in Appendix D,* (α, β) *are identified and* $s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x_{\mathbf{J}}^{(2)}, p_{\mathbf{J}}, \Gamma, F)$ *are identified as functions of* $(\delta_{\mathbf{J}}, x_{\mathbf{J}}^{(2)}, p_{\mathbf{J}}) \in \mathbb{R}^J \times \mathbf{D}_x^{(2)} \times \mathbb{R}^J$.

[33] Common examples of cost-type variables and its proxies are input prices, variables correlated with marginal costs, prices of the same products in other markets (e.g. Hausman-type instruments).

Remark 2. *The two statements of Theorem 2 are complementary: the first statement achieves the identification by restricting the distribution of demand and supply shocks and remains agnostic on the DGP.*[34] *The second statement restricts the DGP and does not posit on the distribution of* $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$.

Proof. See Appendix D. □

The first condition reinforces Assumption 3 to strong exogeneity of $(x_{t\mathbf{J}}, z_{t\mathbf{J}})$ and assumes large support of $z_{t\mathbf{J}}$. The second condition simplifies the price coefficient to be homogeneous for all individuals but still allows for random coefficients on other product characteristics. The third condition imposes the uniqueness of the Bertrand price competition in the factual scenario. The fourth condition is technical and can be further relaxed to allow for "thin tail" distributions F.

Relying on the same data availability, the main result of Allen and Rehbeck (2019a) implies the identification of product-level market share functions in the context of model (4) with additive separable unobservable heterogeneity. While their identification strategy crucially relies on the assumption of additively separable unobservable heterogeneity and does not allow for endogenous prices, I exploit exogenous variation in cost shifters and product characteristics to deal with endogenous prices and market shares, achieving the identification of product-level market share functions.

4.3 Identification of Bundle-Level Market Share Functions

In this section, I assume that product-level market share functions are identified and aim to separately identify Γ and F.[35] To simplify the notation, I include $p_{t\mathbf{J}}$ in $x_{tj}^{(2)}$. The task of separable identification s challenging because only product-level (rather than bundle-level) market shares are available. First,I provide an identification result for a class of mixed-logit models of demand for bundles widely used in the empirical literature.[36]

Theorem 3. *Suppose that* $\mathbf{C}_2 = \{(j, j') : j < j', j, j' \in \mathbf{J}\}$*, or* $\mathbf{C}_2 = \{(j_1, j_2) : j_1 \in \mathbf{J}_1, j_2 \in \mathbf{J}_2\}$*,* $\Gamma_{it\mathbf{b}} = \Gamma_{\mathbf{b}}$ *for* $\mathbf{b} \in \mathbf{C}_2$*, and* $s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ *is identified for* $(\delta_{\mathbf{J}}, x_{t\mathbf{J}}^{(2)}) \in \mathbb{R}^J \times \mathbf{D}_x^{(2)}$*. Then,*

- Γ *and* $s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, F)$ *are identified in* $\mathbb{R}^{C_1} \times \mathbf{D}_x^{(2)}$*, for any* $\mathbf{b} \in \mathbf{C}_1$*.*

[34] The identification in Theorem 2 can also be achieved when the distribution of $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ has "fat tail". See Mattner (1992) and D'Haultfoeuille (2011) for details.

[35] As shown in the second statement of Proposition 1, one can also aim to identify Γ and $s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, F)$ when $\alpha_i = \alpha$.

[36] See Gentzkow (2007), Fan (2013), Kwak et al. (2015), Grzybowski and Verboven (2016).

- *If $\mathbf{D}_x^{(2)}$ is open, and $(\Delta\beta_i^{(2)}, \Delta\alpha_i)$ and $(\eta_{ij})_j^J$ are mutually independent given $x_{t\mathbf{J}}$, then Γ and F are identified.*

Proof. See Appendix E. □

Remark 3. *If for some bundle* $\mathbf{b}$*, the true* $\Gamma_{\mathbf{b}}$ *is equal to* $-\infty$*, i.e. bundle* $\mathbf{b}$ *is not in the choice set, then Theorem 3 implies that* $\Gamma_{\mathbf{b}} = -\infty$ *is identified.*

The first statement of Theorem 3 achieves the separable identification of Γ and $s_{\mathbf{b}}(\delta_t; x_{t\mathbf{J}}^{(2)}, F)$. The second statement achieves the separable identification Γ and F under a mild support on $x_{t\mathbf{J}}^{(2)}$ and the conditional independence between random slopes $(\Delta\beta_i^{(2)}, \Delta\alpha_i)$ and random intercepts $(\eta_{ij})_{j=1}^J$. This independence condition is employed in many empirical papers.[37]

Theorem 3 shows that observing demand data at product level already suffices to identify bundle-level demand primitives in models of demand for multiple products within/across categories. Consequently, researchers are able to conduct nonlinear pricing counterfactuals using these models (see Proposition 1). However, the separable identification in Theorem 3 may not be achieved in some other types of model (4). The following corollary gives an example.

Corollary 1 (Non-separable identification of Γ and F)**.** *Suppose that the data generating process is a model of demand for multiple units:* $\mathbf{J} = \{1\}$ *and* $\mathbf{C}_2 = \{(1,1)\}$*,* $\Gamma_{i(1,1)} = \Gamma > -\infty$*. Moreover, product-level market share function*

$$s_{1\cdot}(\delta; \Gamma, F) = \int \frac{e^{\delta+\mu} + 2e^{2\delta+2\mu+\Gamma}}{1 + e^{\delta+\mu} + e^{2\delta+2\mu+\Gamma}} dF(\mu). \tag{9}$$

is identified. Then, there exists (Γ, F)*, such that* Γ *and* F *are not separably identified.*

Proof. See Appendix F. □

Corollary 1 illustrates the limited power of product-level market shares in models of demand for multiple units to separably identify Γ and F. Intuitively, one cannot distinguish Γ and F because it is impossible to shift the mean utility of the first unit without shifting that of the second unit. When bundle-level demand data is available, Iaria and Wang (2019) shows how to identify and estimate model of demand for bundles by exploring the same bundle-specific fixed effects $\Gamma_{\mathbf{b}}$ across markets. This gives rise to additional moment restrictions that separately identify

[37] In particular, when the random intercepts are degenerated (see Nevo (2000, 2001), Petrin (2002), Berto Villas-Boas (2007), Fan (2013) among many others), or the random coefficients (rather than the random intercepts) are degenerated (see Gentzkow (2007) for example), this condition is automatically satisfied.

Γ and F. With only product-level demand data, this source of identification is no longer available in model (9). Then, unless imposing additional assumptions on synergy parameters or the distribution of random coefficients, the availability of bundle-level demand data may be necessary to disentangle Γ and F and to conduct nonlinear pricing counterfactuals in models of demand for multiple units.[38]

5 Estimation and Implementation

In this section, I propose a GMM estimation procedure for model (4) and discuss its implementation. The proposed estimation procedure is conceptually similar to that used in BLP models of single products. However, due to multiple purchases, the implementation has non-trivial challenges. I consider parametric estimation of model (4) and F is characterised by $\Sigma \in \Theta_\Sigma \subset \mathbb{R}^P$. Define the true value of parameter vector as $\theta_0 = (\alpha_0, \beta_0, \Sigma_0, \Gamma_0)$. Suppose that $(x_{t\mathbf{J}}, z_{t\mathbf{J}})$ are valid instruments and θ_0 is identified.

5.1 Estimation Procedure

I construct unconditional moment conditions from (D.1) using a finite set of functions of $(x_{t\mathbf{J}}, z_{t\mathbf{J}})$, $\Psi = \{\phi_g(x_{t\mathbf{J}}, z_{t\mathbf{J}})\}_{g=1}^G$:

$$m(\theta'; \{\jmath_{t\mathbf{J}.}, p_{t\mathbf{J}}, x_{t\mathbf{J}}, z_{t\mathbf{J}}\}_{t=1}^T, \Psi) = \left(\mathbb{E}\left[\xi_j(\beta', \alpha', \Gamma', F'; \jmath_{t\mathbf{J}.}, x_{t\mathbf{J}}, p_{t\mathbf{J}})\phi_g(z_{t\mathbf{J}}, x_{t\mathbf{J}})\right]\right)_{g=1}^G,$$

The finite-sample counterparts are:

$$m_T(\theta'; \{\jmath_{t\mathbf{J}.}, p_{t\mathbf{J}}, x_{t\mathbf{J}}, z_{t\mathbf{J}}\}_{t=1}^T, \Psi) = \left(\frac{1}{T}\sum_{t=1}^T \frac{1}{J}\sum_{j=1}^J \left[s_j^{-1}(\jmath_{t\mathbf{J}.}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma', \Sigma') - x_{tj}\beta' + \alpha' p_{tj}\right]\phi_g(x_{tj}, z_{tj})\right)_{g=1}^G.$$

Then, the GMM estimator of θ_0, $\hat{\theta}_T^{GMM}$, is defined as:

$$\hat{\theta}_T^{GMM} = \underset{\theta' \in \Theta}{\operatorname{argmin}}\ m_T(\theta'; \{\jmath_{t\mathbf{J}.}, p_{t\mathbf{J}}, x_{t\mathbf{J}}, z_{t\mathbf{J}}\}_{t=1}^T, \Psi)^{\mathrm{T}} W_T m_T(\theta'; \{\jmath_{t\mathbf{J}.}, p_{t\mathbf{J}}, x_{t\mathbf{J}}, z_{t\mathbf{J}}\}_{t=1}^T, \Psi), \tag{10}$$

where Θ is a compact set and $W_T \in \mathbb{R}^{G\times G}$ is a weighting matrix that converges in probability to a positive-definite matrix W. If θ_0 lies in the interior of Θ, then under standard regularity conditions (see Newey and McFadden (1994)), $\hat{\theta}_T^{GMM}$ is consistent and asymptotically normal.[39]

[38] For other papers on the identification with bundle-level demand data, see Fox and Lazzati (2017), Allen and Rehbeck (2019b).

[39] If some parameters (e.g. distributional parameters Σ) are on the boundary, the GMM estimator may not be asymptotically normal. See Ketz (2019) for an inference procedure that is valid when distributional parameters are on the boundary and Andrews (2002) for a general treatment.

A basic requirement for the good finite-sample performance of (10) is that we have at least as many moment conditions as the dimension of $(\alpha_0, \beta_0, \Gamma_0, \Sigma_0)$. In particular, we have $\dim(\Gamma_0)$ demand synergy parameters in (10) that BLP models of single products do not have. Therefore, we need at least $\dim(\Gamma_0)$ more moment conditions. If the number of valid instruments or variability of these instruments is limited, one can also specify Γ_0 and reduce its dimensionality according to the economic setting. For example, two products of the same producer (products of the same store), or of complementary ingredients (flavoured RTE cereals with unflavoured milk), may have greater demand synergies. Then, one can specify demand synergy as a function of product characteristics in a bundle.

In BLP models of single products, a suggested practice is to approximate the optimal instruments in the form of Amemiya (1977) and Chamberlain (1987) that achieve the semi-parametric efficiency bound. Reynaert and Verboven (2014) and Conlon and Gortmaker (2020) report significant gain when impelenting Berry et al. (1995)'s GMM estimator using optimal instruments. However, the difficulty of approximating optimal instruments still remains in estimation procedure (10). A good approximation of optimal instruments relies on the knowledge of the true parameters. Moreover, when the number of products is large, even low order of such approximation may be subject to a curse of dimensionality and the number of needed basis functions is exponentially proportional to the number of products. Gandhi and Houde (2019) provide a solution that breaks the dependence of basis functions on product identity under symmetry conditions among products. The number of basis functions is then invariant with respect to the number of products. However, due to potentially heterogeneous synergy parameters across bundles, product identity matters in (10). The extension of their method to models of demand for bundles is beyond the scope of this paper and I leave it as future research.

5.2 Implementation of Demand Inverse

A key step of the estimation procedure is implementation of the demand inverse in Theorem 1. Given $(x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma', F')$, it seeks for the solution of the following equation:

$$s_{\mathbf{J}\cdot}(\delta'_{t\mathbf{J}}; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma', F') - \mathscr{s}_{t\mathbf{J}\cdot} = 0. \tag{11}$$

In most cases, $s_{\mathbf{J}\cdot}(\cdot; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma', F')$ does not have an analytic form. In practice, researchers often use Monte Carlo method to approximate $s_{\mathbf{J}\cdot}(\cdot; x_{t\mathbf{J}}^{(2)}, p_{t\mathbf{J}}, \Gamma', F')$. In this method, one first draws a finite set of random numbers and then use these fixed random numbers to approximate F'.[40] As a result, F' is numerically implemented

[40]A typical method is to simulate a fixed set of random numbers from uniform distribution in $[0, 1]$ and use $(F')^{-1}$ to transform these random numbers to those under distribution F'.

by a discrete distribution with finite (and therefore compact) support. For this reason, I will assume that F' has compact support when analysing the numerical performance of the implementation of the demand inverse (11).

In models of demand for single products, Berry et al. (1995) propose a fixed-point iterative algorithm to implement the demand inverse. An essential property of this algorithm is contraction mapping, which guarantees the convergence of the iteration. However, the contraction-mapping property may not hold if one uses the same iterative algorithm to solve (11) because products can be Hicksian complementary. To solve this challenge, I propose to use Jacobian-based approach to solve (11). This approach has been adopted in the literature. Conlon and Gortmaker (2020) tests performances of different Jacobian-based algorithms to implement the demand inverse in models demand for single products, and find supportive evidences for the numerical efficiency of Jacobian-based methods. A leading example is Newton-Raphson method:

$$\begin{aligned} &\delta^{(0)} = \delta^{(0)}, \\ &\delta^{(n+1)} = \delta^{(n)} - J_s^{-1}(\delta^{(n)})[s_{\mathbf{J}.}(\delta'_{t\mathbf{J}}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma', F') - \mathscr{s}_{t\mathbf{J}.}], \end{aligned} \tag{12}$$

where $J_s(\delta'_{t\mathbf{J}}) = \frac{\partial s_{\mathbf{J}.}(\delta'_{t\mathbf{J}}; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \Gamma', F')}{\partial \delta'_{t\mathbf{J}}}$. To solve (11), Algorithm (12) is well-defined because $J_s(\delta'_{t\mathbf{J}})$ is everywhere symmetric and positive-definite. Moreover, the uniqueness of solution is guaranteed by Theorem 1: if (12) converges, then it must converge to the unique solution of (11).

It is well-known that the numerical performance of Jacobian-based algorithms such as (12) crucially depends on the quality of starting value $\delta^{(0)}$: the closer $\delta^{(0)}$ is to the solution $\delta'_{t\mathbf{J}}$, the faster Algorithm (12) converges.[41] In the setting of (11), we can leverage econometric properties of the demand model to construct such a starting value that is directly constructed from data and "close" to the solution of (11). The next proposition gives an example:

Proposition 2. *Suppose that* $\mathscr{s}_{t\mathbf{J}.}$ *in* (11) *are generated from a model of demand for multiple products across* K *categories, for* $K \geq 1$ *(see Section 3.1) and the distribution* F' *has compact support* $\mathbf{D}_F$*. Denote the solution to* (11) *by* $\delta'_{\mathbf{J}}$*. For products of category* k*, define* $\delta^{(0)}_{k*} = \left(\delta^{(0)}_{jk*}\right)_{j\in\mathbf{J}_k}$*, where* $\delta^{(0)}_{jk*} = \ln \frac{\mathscr{s}_{tj.}}{1-\sum_{j\in\mathbf{J}_k} \mathscr{s}_{tj.}}$*, and* $\delta^{(0)}_{*} = (\delta^{(0)}_{k*})_{k=1}^{K}$*. Then, there exists a constant* $A(\mathbf{D}_F, \Gamma') > 0$ *such that*

$$|\delta'_{\mathbf{J}} - \delta^{(0)}_{*}| \leq A(\mathbf{D}_F, \Gamma').$$

Proof. See Appendix G. □

[41] One basic result on global convergence of Newton-Raphson method in numerical analysis is Newton-Kantorovich Theorem. See Ortega (1968) for the statement and one proof.

Even though it is hard to derive similar results in a general model (4), Proposition 2 sheds light on how to find a good starting value for Jacobian-based algorithms: it suggests to use a starting point as if the DGP is a multinomial logit model. In a model with the bundle size being up to size K, an starting value along the lines of Proposition 2 can be defined as

$$\delta_{j*}^{(0)} = \ln \frac{\mathcal{s}_{j\cdot}}{K - \sum_{j \in \mathbf{J}} \mathcal{s}_{j\cdot}}. \tag{13}$$

for $j \in \mathbf{J}$. Here $K - \sum_{j \in \mathbf{J}} \mathcal{s}_{j\cdot}$ serves as the "market share" of the outside option.[42] In Appendix K, I report numerical gains of using $\delta_*^{(0)}$ in Monte Carlo simulations.

6 Empirical Illustration

In this section, I illustrate the practical implementation of the proposed methods and estimate the demand for Ready-To-Eat (RTE) cereals and milk in the US. In particular, I demonstrate the economic importance of having (flexible) deman synergies between RTE cereal and milk products in demand.

To do so, I first estimate three models of demand for bundles: model I in which demand synergy parameters $\Gamma_{\mathbf{b}} = 0$ for all $\mathbf{b}$, model II in which $\Gamma_{\mathbf{b}} = \gamma_0$ for any $\mathbf{b}$ but not necessarily zero (it will be estimated), and a full model in which $\Gamma_{\mathbf{b}}$'s vary across bundles as functions of product characteristics in the bundle. In model I, the demand for RTE cereals and that for milk are restricted to be independent for each household (the cross-price elasticities between RTE cereals and milk are hence zero); in model II, the single synergy parameter γ_0 captures the prevalent synergy in consumption of RTE cereals and milk, but not the potential heterogeneity in synergies across different bundles; the full model allows such synergies to depend on the composition of a bundle, i.e. how well RTE cereal and milk characteristics are matched. I simulate a merger between a major RTE cereal producer and a milk producer using each model. Comparisons of the merger simulations across the three models show that ignoring the synergies (model I), or restricting them to be uniform (model II) may result in important bias in welfare prediction.

6.1 Data and Definitions

I use the store-week level datasets of the RTE cereal and milk categories from the IRI data. The IRI data has been used in the empirical literature of demand (see

[42] When the bundle size is up to K, since product-level market shares of two different products overlap on up to $K-1$ bundle-level market shares. Consequently, the sum of all product-level market shares is strictly smaller than K.

Nevo (2000, 2001)). I will give a succinct description and refer to these papers and also Bronnenberg et al. (2008) for a thorough discussion.

I focus on the period 2008-2011 and the city of Pittsfield in the US. I define a market t as a combination of store and week and obtain 1387 markets. In each market, the sales (in lbs and dollars) of RTE cereals and fluid milk are observed at Universal Product Code (UPC) level. For the RTE cereal category, similarly to Nevo (2001), I define a product as a combination of brand, flavour, fortification, and type of grain. For fluid milk category, I define a product as a combination of brand, flavour, fortification, fat content, and type of milk. Then, the sales of product j of category $k \in \mathbf{K} = \{RTE\ cereal, fluid\ milk\}$ in market t is the sum of the sales in lbs of all the UPC's that this product collects. The price of j of category k in market t, p_{tj}^k, is defined as the ratio between its sales in dollars and in lbs. I define the choice set of RTE cereals as that of the 25 largest RTE cereal products (in terms of sales in lbs), collecting all the other smaller products in the outside option for cereal category. Similarly, I define the choice set of milk as that of the 20 largest fluid milk products.[43] Denote the choice set by $\mathbf{J}_k$ for $k \in \mathbf{K}$.

For each market, I consider the weekly consumption of breakfast cereals as the market size for RTE cereal category, and weekly consumption of fluid milk for milk category. To calibrate the market size for each category, I assume that households go shopping once per week for breakfast cereals and fluid milk. Then, the market size for RTE cereal category (or milk) is the product of the weekly per capita consumption of breakfast cereals (or fluid milk) and the sampled population size. I obtain the former information from external sources and the latter from the IRI data. Finally, for each market, the product-level market share of $j \in \mathbf{J}_k$ is then the ratio between its total sales in lbs and the market size for category k. Appendix H provides computational details of the construction of the product-level market shares. Tables 9-10 in Appendix I summarise product characteristics.

6.2 Model Specification

For each store-week combination t, denote the set of available products in category $k \in \mathbf{K}$ by $\mathbf{J}_{tk} \subset \mathbf{J}_k$. Denote by 1 the RTE cereal category, 2 the milk category, and $\mathbf{J}_t = \mathbf{J}_{t1} \cup \mathbf{J}_{t2}$. The set of bundles $\mathbf{C}_{t2}$ is defined as $\mathbf{J}_{t1} \times \mathbf{J}_{t2}$, where each bundle contains a RTE cereal product and a milk product.[44] Household's choice set is then defined as $\mathbf{C}_t = \mathbf{J}_t \cup \mathbf{C}_{t2} \cup \{0\}$, where 0 represents the outside option.[45] The size

[43]The purchase of the 25 RTE cereal products represents 38% of the total purchase of RTE cereals in the IRI data, and that of the 20 fluid milk represents around 88%.

[44]I do not include bundles of products of the same category.

[45]According to the definition of products and the market sizes, the outside option collects RTE cereals or milk products that are not included in $\mathbf{J}_k$, relevant products not present in the categories (e.g. cereal biscuits), and the bundles of these products (e.g. cereal biscuits and milk).

of $\mathbf{C}_t$ is up to 546 (45 products, 500 bundles, 1 outside option) if all products in $\mathbf{J}$ are available in market t.

For household i in market t, the indirect utility from purchasing product $j \in \mathbf{J}_{tk}$ is:

$$\begin{aligned}
U_{itj}^k &= -p_{tj}^k \alpha_i + \eta_{ij}^k + \xi_{tj}^k + \varepsilon_{itj}^k \\
&= [-p_{tj}^k \alpha + \eta_j^k + \xi_{tj}^k] + [\Delta\eta_{ij}^k - \Delta\alpha_i p_{tj}^k] + \varepsilon_{itj}^k \\
&= \delta_{tj}^k + \mu_{itj}^k + \varepsilon_{itj}^k, \\
\mu_{itj}^k &= \Delta\eta_{ij}^k - (d_i \Delta\alpha + v_i) p_{tj}^k,
\end{aligned}$$

and

$$\Delta\eta_{ij}^k = \begin{cases} \Delta\eta_{i,\text{flavor}(j)} + \Delta\eta_{i,\text{fortification}(j)} + \Delta\eta_{i,\text{brand}(j)}^1, & \text{if } k = 1, \\ \Delta\eta_{i,\text{flavor}(j)} + \Delta\eta_{i,\text{fortification}(j)} + \Delta\eta_{i,\text{brand}(j)}^2 + \Delta\eta_{i,\text{fat content}(j)}^2, & \text{if } k = 2, \end{cases}$$

where δ_{tj}^k is market t-specific mean utility for $j \in \mathbf{J}_{tk}$, μ_{itj}^k is a household i-specific utility deviation from δ_{tj}^k, and ε_{itj}^k is an idiosyncratic error term. α is population-average price coefficient, and $\Delta\alpha_i = d_i \Delta\alpha + v_i$ is household i-specific price coefficient deviation from α and is the sum of an observed part that is a function of the household characteristics d_i (income groups) and an unobserved component v_i. $\Delta\eta_{ij}^k$ is an unobserved household i-specific preference for product j of category k, where $\Delta\eta_{i,\text{flavor}(j)}$ captures household i's unobserved preference for the flavour of j of category k (unflavoured, flavoured), $\Delta\eta_{i,\text{fortification}(j)}$ captures i's unobserved preference for the nutrition in product j of category k (unfortified, fortified), $\Delta\eta_{i,\text{brand}(j)}^k$ captures i's unobserved preference for the brand of j of category k, and $\Delta\eta_{i,\text{fat content}(j)}^2$ captures i's unobserved preference for the fat content in milk j (whole fat, low fat, skimmed). Note that because RTE cereals and milk have both flavour and fortification characteristics. Then, for products $j \in \mathbf{J}_1$ and $r \in \mathbf{J}_2$, if they have the same flavour (or fortification type), then $\Delta\eta_{i,\text{flavor}(j)} = \Delta\eta_{i,\text{flavor}(r)}$ (or $\Delta\eta_{i,\text{fortification}(j)} = \Delta\eta_{i,\text{fortification}(r)}$).

The indirect utility of household i in market t from purchasing bundle $\mathbf{b} = (j, r)$ is:

$$\begin{aligned}
U_{it\mathbf{b}} &= \left[-p_{tj}^1 \alpha_i + \eta_{ij}^1 + \xi_{tj}^1\right] + \left[-p_{tr}^2 \alpha_i + \eta_{ir}^2 + \xi_{tr}^2\right] + \Gamma_{\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\
&= \left[\delta_{tj}^1 + \delta_{tr}^2 + \Gamma_{\mathbf{b}}\right] + \left[\mu_{itj}^1 + \mu_{itr}^2\right] + \varepsilon_{it\mathbf{b}} \\
&= \delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}) + \mu_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}},
\end{aligned}$$

where $\delta_{t\mathbf{b}}(\Gamma_{\mathbf{b}}) = \delta_{tj}^1 + \delta_{tr}^2 + \Gamma_{\mathbf{b}}$ is market t-specific mean utility for bundle $\mathbf{b}$, $\mu_{it\mathbf{b}}$ is household i-specific utility deviation from $\delta_{t\mathbf{b}}$, $\Gamma_{\mathbf{b}} = \Gamma_{(j,r)}$ is demand synergy between RTE cereal j and milk r, and $\varepsilon_{it\mathbf{b}}$ is an idiosyncratic error term. Demand synergy parameter $\Gamma_{(j,r)}$ captures the extra utility household obtains from buying

RTE cereal j and milk r jointly rather than separately. One prominent reason for the joint purchase is synergy in consumption, i.e. members in the household consume together RTE cereals and milk for their breakfasts.[46] The quality of match between the characteristics of RTE cereal j and milk r may determines the extra utility $\Gamma_{(j,r)}$. Consequently, I specify $\Gamma_{(j,r)}$ as a function of the characteristics of j and r:

$$\begin{aligned}\Gamma_{(j,r)}(\gamma) = \gamma_0 &+ \mathbf{1}\{\text{j is multi-grain}\}\gamma_1 + \mathbf{1}\{\text{j is granola}\}\gamma_2 \\ &+ \mathbf{1}\{\text{r is skimmed}\}\gamma_3 + \mathbf{1}\{\text{r is low fat}\}\gamma_4 \\ &+ \mathbf{1}\{\text{j is flavoured}\}\gamma_5 + \mathbf{1}\{\text{r is chocolate milk}\}\gamma_6 \\ &+ \mathbf{1}\{\text{j is flavored and r is chocolate milk}\}\gamma_7 \\ &+ \mathbf{1}\{\text{j is fortified, r is chocolate milk}\}\gamma_8 + \mathbf{1}\{\text{j is fortified, r is fortified}\}\gamma_9.\end{aligned} \tag{14}$$

Parameter γ_0 represents the synergy in consumption of the reference bundle (unflavoured unfortified uni-grain RTE cereal and unflavoured whole-fat milk).γ_1 and γ_2 quantify additional synergies due to other types of grains (multi-grain, granola). γ_3 and γ_4 measures additional synergies due to lower fat content (skimmed, low fat). γ_5, γ_6 and γ_7 proxy additional synergies due to flavour combinations. γ_8 and γ_9 quantify additional synergies due to combinations of fortified nutrition in RTE cereals and milk products.

Finally, the indirect utility of household i in market t from purchasing the outside option is normalised to be $U_{it0} = \varepsilon_{it0}$. Denote the random coefficients by

$$\begin{aligned}\theta_{it} = (v_i, &\Delta\eta_{i,\text{unflavoured}}, \Delta\eta_{i,\text{flavoured}}, \Delta\eta_{i,\text{unfortified}}, \Delta\eta_{i,\text{fortified}}, \\ &\{\Delta\eta^1_{i,\text{br.}}\}_{\text{br.}\in\mathbf{B}_1}, \{\Delta\eta^2_{i,\text{br.}}\}_{\text{br.}\in\mathbf{B}_2}, \Delta\eta^2_{i,\text{whole fat}}, \Delta\eta^2_{i,\text{low fat}}, \Delta\eta^2_{i,\text{skimmed}}),\end{aligned}$$

where $\mathbf{B}_k$ denotes the set of brands in category k. I assume that θ_{it} follows a centred Gaussian distribution F and the components are uncorrelated. Note that this specification already allows for unobserved correlation among products of the same characteristics within and across categories (e.g. same flavour). Define $\delta_{t\mathbf{J}_t} = (\delta^1_{t\mathbf{J}_{t1}}, \delta^2_{t\mathbf{J}_{t2}})$ and $p_{t\mathbf{J}_t} = (p^1_{t\mathbf{J}_{t1}}, p^2_{t\mathbf{J}_{t2}})$. Write $\mu^k_{itj} = \mu^k_{tj}(d_i, \theta_{it}, p^k_{tj})$. Finally, assume that ε_{it0}, ε^1_{itj}'s, ε^2_{itj}'s, and $\varepsilon_{it\mathbf{b}}$'s are i.i.d. Gumbel. Then, the product-level market

[46] Another reason can be shopping cost, i.e. household may not want to go shopping twice to buy RTE cereals and milk. However, this is ruled out in this application because the purchases are made within the same store.

share function of $j \in \mathbf{J}_{t1}$ in market t is:

$$
\begin{aligned}
&s_{j.|\mathbf{J}_t}(\delta_{t\mathbf{J}_t}; p_{t\mathbf{J}_t}, \gamma, F) \\
&= \int \frac{e^{\delta^1_{tj}+\mu^1_{tj}(d_i,\theta_{it},p^1_{tj})}\left[1+\sum_{r\in\mathbf{J}_{t2}} e^{\delta^2_{tr}+\mu^2_{tr}(d_i,\theta_{it},p^2_{tr})+\Gamma_{(j,r)}(\gamma)}\right]}{1+\sum_{k=1,2}\sum_{j'\in\mathbf{J}_{tk}} e^{\delta^k_{tj'}+\mu^k_{tj'}(d_i,\theta_{it},p^k_{tj'})}+\sum_{(j',r)\in\mathbf{J}_{t1}\times\mathbf{J}_{t2}} e^{\delta^1_{tj'}+\delta^2_{tr}+\mu^1_{tj'}(d_i,\theta_{it},p^1_{tj'})+\mu^2_{tr}(d_i,\theta_{it},p^2_{tr})+\Gamma_{(j',r)}(\gamma)}} dF(\theta_{it})d\Pi_t(d_i),
\end{aligned}
\tag{15}
$$

where $\Pi_t(\cdot)$ is the distribution function of demographics d_i in market t. The formula for $r \in \mathbf{J}_{t2}$ is similar.

6.3 Demand Estimates

Table 3 summarises demand estimates. In column "IV regression", I estimate a multinomial logit with $\Gamma_{\mathbf{b}} = 0$ for all $\mathbf{b} \in \mathbf{C}_2$. Columns "Model I", "Model II", and "Full Model" show the estimates by using model I, model II, and the full model, respectively. In all the models, I control for product-specific intercepts and use the same Hausman-type instruments. These instruments include prices of the same products in the same store and week but in other cities (Boston for RTE cereals and Hartford for milk), the prices of other products of the same category with the same product characteristics.

Price coefficient (α) is estimated -0.59 in the multinomial logit model. The estimates of the other three models with random coefficients show important heterogeneity in price sensitivities across income groups. Without surprise, households with higher income are estimated to have a lower (in absolute value) price coefficient and therefore less sensitive to price change. The standard variance of the unobserved heterogeneity in the price coefficient (σ_v) is estimated small. Moreover, after controlling for the product-specific intercepts, households' preference seems to be almost homogeneous for products within some types (e.g. unflavoured, fortification, fat content). One potential reason is that products are little differentiated within each of these types. In contrast, households' preference for flavoured products seems to be more heterogeneous. This is also intuitive because flavours of RTE cereals and milk are much more horizontally differentiated and different households may have their favourite flavours. I also find that households' preference for RTE cereal brands is more heterogeneous than that for milk brands.

In model II, synergy parameter γ_0 is estimated to be 0.902 and significant. This specification constraints all bundles of RTE cereal and milk to have the same synergy, regardless of their characteristics. In the full model, the demand synergies are allowed to vary across bundles. In column "Full Model", $\gamma_0 = -1.540$ represents the synergy in consumption between the unflavoured unfortified uni-grain RTE cereal and unflavoured whole-fat milk. Regarding the characteristics of RTE cereals,

γ_1 is estimated positive, meaning that multi-grain cereals are preferred (over uni-grain ones) when consumed with milk. Moreover, granola, which contains oats and other whole grains as well as ingredients such as dried fruit and nuts, is estimated to be even preferred over multi-grain cereals ($\gamma_2 > \gamma_1$). For the characteristics of milk, products with lower fat are estimated to be preferred when consumed with cereals ($\gamma_3 > \gamma_4 > 0$). Another interesting finding is households' preference for flavour combinations of cereals and milk. Flavoured cereals are estimated to be preferred over unflavoured ones ($\gamma_5 > 0$). For chocolate milk, households' preference seems to be more complicated. When consumed with unflavoured cereals, chocolate milk is preferred over unflavoured milk ($\gamma_6 > 0$). While, I find that it is seldom consumed with flavoured or fortified cereals, i.e. γ_7 and γ_8 are estimated very negative.[47] One potential reason is that flavoured (or fortified) RTE cereals are more likely frosted. Very negative γ_7 and γ_8 may reflect households' disutility for bundles with too much sugar. Finally, I also find that bundles of fortified cereals and milk are less appealing than the reference one ($\gamma_9 < 0$). In the data, the types of added nutrition in RTE cereals and milk are the same, e.g. vitamins, calcium. $\gamma_9 < 0$ may reflect that the same types of added nutrition in cereals and milk are substitute.

6.4 Price Elasticities

I compute the average (across markets) estimated self- and cross-price elasticities obtained from the full model. Because there is no income effect in the specification, negative (positive) cross-price elasticities are then interpreted as Hicksian complementarity (substitutability). To facilitate the exposition, I report the price elasticities at the level of product characteristics and producers. This illustrates how RTE cereals and milk are complementary along each of these dimensions. The results are illustrated in Tables 4-7. Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row producer (or characteristics) with respect to a 1% increase in the prices of the products collected by the column producer (or characteristics).[48]

Overall, RTE cereals are estimated to have larger self-price elasticities than

[47] In Table 3, the estimates of γ_7 and γ_8 are $-\infty$. This means that the model with $\gamma_7, \gamma_8 = -\infty$, i.e. the corresponding bundles are not in the choice set, performs statistically as well as the one without these restrictions in terms of the value of GMM objective function. Concretely, I first estimate a model with γ_7 and γ_8 being finite. I find that the estimated γ_7 and γ_8 are very negative. Then, I estimate the model with $\gamma_7, \gamma_8 = -\infty$. The difference in the value of the two GMM objective functions is less than 10^{-9}, or equivalently, one cannot reject the "hypothesis" that $\gamma_7, \gamma_8 = -\infty$. In the future, I will consider a formal testing procedure.

[48] Concretely, denote by $\mathbf{C}_m$ the set of products that row (column) producer m represents. Then, the price elasticity between brands m and n, E_{mn}, is defined as $E_{mn} = \frac{\sum_{j \in \mathbf{J}_m} s_{j\cdot} \sum_{r \in \mathbf{J}_n} \epsilon_{jr}}{\sum_{j \in \mathbf{J}_m} s_{j\cdot}}$, where ϵ_{jr} is the price elasticity between products j and r.

milk. This may reflect that households view milk more necessary than RTE cereals and therefore are less sensitive to price change of milk. Given the specification of model (15), RTE cereals are always substitutes to each other and the cross-price elasticities among them are positive. Similarly, the cross-price elasticities among milk products are also positive. These are shown by the positive off-diagonal elements in the diagonal blocks (RTE cereals-RTE cereals, Milk-Milk) of Tables 4-7. Differently, the cross-price elasticities between RTE cereal and milk products, i.e. the elements in the off-diagonal blocks (RTE cereals-Milk, Milk-RTE cereals), can be either positive or negative.

Table 4 shows the substitution patterns along the dimensions of grain type and fat content. Granola is estimated to be complementary to milk with any level of fat and skimmed milk is complementary to cereals with any kind of grain. Moreover, milk with lower fat content is uniformly more complementary to any kind of grain than milk with higher fat content. This reveals that households do not seek for fat in milk when consuming it with RTE cereals. As to flavours (Tables 5-6), unflavoured cereals and flavoured milk (and the reverse) are shown to be complementary. In contrast, flavoured cereals and chocolate milk are estimated to be (strong) substitutes. Coherent with the estimates of γ's in Table 3, the relationship between chocolate milk and RTE cereals is more complicated. Chocolate milk is estimated to be complementary to unflavoured or unfortified cereals. However, it is estimated to be substitute to flavoured or fortified cereals. Finally, I find that most RTE cereals and milk are complementary at producer level.

As a comparison, the demand synergies in model I are constrained to be zero. The cross-price elasticities between RTE cereals and milk are therefore mechanically zero. In model II, all the bundles are restricted to have the same demand synergy which is estimated to be positive (see the column "Model II" of Table 3). I re-do the exercises in Tables 4-7 using the demand estimates from model II (see Tables 11-14 of Appendix I). In contrast to those obtained from the full model, the results show that RTE cereals and milk are complementary along every dimension.

Table 3: Demand Estimates

	IV Regression $\Gamma_{\mathbf{b}} = 0$	Model I $\Gamma_{\mathbf{b}} = 0$	Model II $\Gamma_{\mathbf{b}} = \gamma_0$	Full Model
Price Coef.				
uniform, α	-0.59 (0.011)			
(baseline) low income, α_1		-1.369 (0.042)	-1.128 (0.060)	-1.062 (0.075)
medium income, $\Delta\alpha_2$		0.218 (0.164)	0.148 (0.096)	0.164 (0.0764)
high income, $\Delta\alpha_3$		0.947 (0.045)	0.718 (0.029)	0.712 (0.0269)
Random Coef.				
σ_v		0.115 (0.070)	0.086 (0.072)	0.046 (0.1244)
$\sigma_{\text{unflavoured}}$		0.015 (2.754)	0.018 (1.490)	0.023 (2.6772)
$\sigma_{\text{flavoured}}$		2.353 (0.259)	1.684 (0.038)	1.010 (0.1615)
$\sigma_{\text{unfortified}}$		0.048 (3.047)	0.004 (3.870)	0.017 (2.2994)
$\sigma_{\text{fortified}}$		0.010 (5.393)	0.010 (2.388)	0.015 (6.8129)
σ_{fat}		0.077 (1.099)	0.062 (0.660)	0.034 (1.6161)
$\sigma_{\text{cereal brand}}$		0.780 (0.049)	0.660 (0.058)	0.847 (0.0705)
$\sigma_{\text{milk brand}}$		0.005 (6.156)	0.003 (5.368)	0.004 (6.9505)
Demand Synergies				
γ_0			0.902 (0.155)	-1.540 (0.3437)
multi-grain, γ_1				0.533 (0.0359)
granola, γ_2				4.363 (0.0891)
skimmed, γ_3				2.880 (0.2111)
low fat, γ_4				0.514 (0.1282)
flavoured cereal, γ_5				1.816 (0.2324)
chocolate milk, γ_6				13.625 (0.2621)
flavoured cereal and chocolate milk, γ_7				$-\infty$
fortified cereal, chocolate milk, γ_8				$-\infty$
fortified cereal and milk, γ_9				-1.538 (0.3512)
GMM Objective Function		0.1636	0.1599	0.1434

Notes: Standard errors are reported in brackets. For all the models, instruments are the same and product-specific intercepts are included. In the "IV Regression", week dummies and store dummies are also included.

Table 4: Average Estimated Own- and Cross-Price Elasticities (Full Model), Grain Type and Fat Content

	RTE cereals			Milk		
	uni-grain	multi-grain	granola	skimmed	low fat	whole fat
RTE cereals, uni-grain	−1.407	0.194	0.009	−0.032	0.007	0.009
	(0.221)	(0.049)	(0.002)	(0.007)	(0.003)	(0.003)
multi-grain	0.266	−1.492	0.009	−0.034	0.001	0.009
	(0.063)	(0.221)	(0.011)	(0.007)	(0.003)	(0.003)
granola	0.220	0.168	−1.335	−0.084	−0.071	−0.005
	(0.054)	(0.182)	(0.207)	(0.015)	(0.011)	(0.006)
Milk, skimmed	−0.350	−0.243	−0.053	−0.252	0.047	0.023
	(0.068)	(0.051)	(0.010)	(0.043)	(0.029)	(0.013)
low fat	0.010	−0.005	−0.020	0.018	−0.262	0.028
	(0.012)	(0.007)	(0.003)	(0.012)	(0.040)	(0.018)
whole fat	0.056	0.045	−0.005	0.018	0.054	−0.307
	(0.023)	(0.018)	(0.004)	(0.011)	(0.036)	(0.055)

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row characteristics with respect to a 1% increase in the prices of the products collected by the column characteristics. Standard errors are reported in brackets and computed from a parametric bootstrap as in Nevo (2000, 2001) with 200 draws.

Table 5: Average Estimated Own- and Cross-Price Elasticities (Full Model), Flavours

	RTE cereals		Milk	
	unflavoured	flavoured	unflavoured	chocolate
RTE cereals, unflavoured	−1.397	0.190	0.016	−0.014
	(0.242)	(0.062)	(0.004)	(0.003)
flavoured	0.145	−1.381	−0.051	0.003
	(0.047)	(0.240)	(0.007)	(0.001)
Milk, unflavoured	0.031	−0.130	−0.214	0.001
	(0.007)	(0.017)	(0.029)	(0.001)
chocolate	−1.319	0.378	0.071	−0.264
	(0.273)	(0.141)	(0.049)	(0.045)

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row characteristics with respect to a 1% increase in the prices of the products collected by the column characteristics. Standard errors are reported in brackets and computed from a parametric bootstrap as in Nevo (2000, 2001) with 200 draws.

Table 6: Average Estimated Own- and Cross-Price Elasticities (Full Model), Fortification and Flavours

	RTE cereals		Milk	
	unfortified	fortified	unflavoured	chocolate
RTE cereals, unfortified	−1.263	0.060	−0.029	−0.006
	(0.261)	(0.021)	(0.004)	(0.002)
fortified	0.393	−1.668	0.043	0.002
	(0.132)	(0.349)	(0.009)	(0.000)
Milk, unflavoured	−0.124	0.025	−0.214	0.001
	(0.017)	(0.006)	(0.029)	(0.001)
chocolate	−1.004	0.063	0.071	−0.264
	(0.320)	(0.007)	(0.049)	(0.045)

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row characteristics with respect to a 1% increase in the prices of the products collected by the column characteristics. Standard errors are reported in brackets and computed from a parametric bootstrap as in Nevo (2000, 2001) with 200 draws.

Table 7: Average Estimated Own- and Cross-Price Elasticities (Full Model), Brands

	RTE cereals					Milk				
	General Mills	Kashi	Kellogg's	Post	Private Label	Garelick Farms	Guidas	High Lawn Farm	Hood	Private Label
RTE cereals, General Mills	−1.444 (0.229)	0.009 (0.003)	0.146 (0.034)	0.037 (0.009)	0.009 (0.004)	−0.007 (0.001)	−0.034 (0.005)	−0.005 (0.001)	−0.004 (0.001)	−0.022 (0.003)
Kashi	0.154 (0.044)	−1.678 (0.254)	0.152 (0.176)	0.035 (0.046)	0.011 (0.023)	−0.005 (0.001)	−0.005 (0.001)	−0.0003 (0.000)	−0.0002 (0.000)	−0.001 (0.002)
Kellogg's	0.152 (0.036)	0.010 (0.012)	−1.424 (0.212)	0.032 (0.024)	0.009 (0.012)	−0.002 (0.001)	0.006 (0.001)	−0.0005 (0.000)	−0.0003 (0.000)	0.002 (0.001)
Post	0.173 (0.043)	0.010 (0.016)	0.146 (0.128)	−1.541 (0.210)	0.010 (0.015)	−0.005 (0.001)	−0.013 (0.002)	−0.002 (0.000)	−0.001 (0.000)	−0.009 (0.002)
Private Label	0.175 (0.069)	0.013 (0.029)	0.163 (0.178)	0.038 (0.047)	−1.335 (0.207)	−0.044 (0.006)	.	−0.012 (0.002)	−0.011 (0.002)	−0.094 (0.012)
Milk, Garelick Farms	−0.092 (0.014)	−0.002 (0.001)	−0.018 (0.007)	−0.014 (0.003)	−0.032 (0.005)	−0.323 (0.079)	.	0.005 (0.007)	0.005 (0.006)	0.054 (0.046)
Guidas	−0.044 (0.006)	−0.0003 (0.000)	0.0004 (0.001)	−0.004 (0.001)	.	.	−0.173 (0.024)	0.012 (0.006)	0.002 (0.001)	0.014 (0.006)
High Lawn Farm	−0.109 (0.016)	0.001 (0.000)	−0.006 (0.008)	−0.011 (0.002)	−0.020 (0.004)	0.015 (0.017)	0.119 (0.062)	−0.479 (0.121)	0.004 (0.006)	0.046 (0.041)
Hood	−0.154 (0.023)	−0.001 (0.001)	−0.032 (0.011)	−0.018 (0.004)	−0.030 (0.007)	0.015 (0.017)	0.114 (0.060)	0.007 (0.009)	−0.497 (0.112)	0.045 (0.040)
Private Label	−0.076 (0.011)	0.001 (0.000)	0.011 (0.006)	−0.007 (0.001)	−0.022 (0.003)	0.016 (0.014)	0.123 (0.056)	0.007 (0.006)	0.005 (0.004)	−0.251 (0.045)

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row producer with respect to a 1% increase in the prices of the products collected by the column producer. Standard errors are reported in brackets and computed from a parametric bootstrap as in Nevo (2000, 2001) with 200 draws.
The . refers to the situation where the products of the row and column brands are not available at the same time in any market in the data.

6.5 Counterfactual Simulations

I simulate a merger between a major RTE cereal producer (General Mills) and a milk producer (Garelick Farms). To do so, I assume constant market-product specific marginal production costs and these marginal costs remain unchanged after the merger, i.e. there is no efficiency gain from the merger. Moreover, I assume that in both factual and counterfactual scenarios producers play a simultaneous Bertrand price-setting game with complete information and linear pricing strategies. Then, I can back out these marginal costs from the estimated demand and the observed prices by inverting the First-Order Conditions of the pricing game in the factual scenario. Finally, I simulate the merger using the estimated demand and marginal costs. I replicate this merger simulation using the full model, model I ($\Gamma_{\mathbf{b}} = 0$), and model II ($\Gamma_{\mathbf{b}} = \gamma_0$). Table 8 summarises the results.

Model I restricts all the demand synergies between RTE cereals and milk to be zero. As a result, their cross-price elasticities are always zero and the merger between General Mills and Garelick Farms will not lead to any change in prices and consumer surplus (row "Model I" in Table 8). In contrast, both model II and the full model allow for demand synergies and estimate that RTE cereals and milk products exhibit substantial complementarity at the producer level (Tables 7 and 14). The merged producer internalises the complementarity in the pricing and therefore reduces the prices of General Mills RTE cereals and Garelick Farms milk. As a result, the overall prices in both categories decrease after the merger and consumer surplus increases. This is coherent with Cournot (1838)'s intuition that mergers between producers selling complementary products can be socially desirable.

It is worth noting that model II seems to overestimate the consumer surplus gain, nearly 70.3% more than the full model. This is due to the specification $\Gamma_{\mathbf{b}} = \gamma_0$ that restricts *all* RTE cereals and milk to have the same level of synergy in consumption, regardless of their characteristics, resulting in too much complementarity. In contrast, the full model specifies the synergy as a function of characteristics of RTE cereals and milk, allowing for flexible synergy patterns in consumption. Consequently, as shown in Tables 4-6, the cross-price elasticities between RTE cereals and milk depend crucially on the match of their characteristics (flavours, grain type, fat content, etc.), some pairs creating less synergy than others (e.g. flavoured RTE cereal and milk in Table 5). Even though both model II and the full model capture the first-order effect of complementarity between RTE cereals and milk, the cross-price elasticities and this merger exercise suggest that it is plausible in empirical research to use a model of demand for bundles that allows for flexible synergy patterns in demand.

Table 8: Merger Simulation, General Mills and Garelick Farms

	Price change		Consumer Surplus
	RTE Cereals	Milk	change
Model I, $\Gamma_{\mathbf{b}} = 0$	0%	0%	0%
Model II, $\Gamma_{\mathbf{b}} = \gamma_0$	−0.49%	−4.49%	3.44%
Full Model	−0.67%	−3.49%	2.02%

Notes: The Table reports average price changes (first two columns) and consumer surplus change (last column) after the merger between General Mills and Garelick Farms, with respect to the observed oligopoly. The first row refers to the model of demand imposing $\Gamma_{\mathbf{b}} = 0$ for any $\mathbf{b}$ in demand estimation (column "Model I" in Table 3). The second row refers to model II that restricts all $\Gamma_{\mathbf{b}} = \gamma_0$ in estimation (column "Model II" in Table 3). The third row refers to the full model model II (column "Full Model" in Table 3). The counterfactual is simulated for markets in which all RTE cereal products and private label products are available.

7 Conclusion

This paper considers the identification and estimation of a random coefficient discrete choice model of bundles, namely sets of products, when only product-level market shares are available. This last feature arises when only aggregate purchases of products, as opposed to individual purchases of bundles, are available, a very common phenomenon in practice. Following the classical approach with aggregate data, I consider a two-step method. First, using a novel inversion result in which demand can exhibit Hicksian complementarity, I recover the mean utilities of products from product-level market shares. Second, to infer the structural parameters from the mean utilities while dealing with price endogeneity, I use IVs. I propose a practically useful GMM estimator whose implementation is straightforward, essentially as a standard BLP estimator. Finally, I estimate the demand for RTE cereals and milk in the US. The demand estimates suggest that RTE cereals and milk are overall complementary and the synergy in consumption crucially depends on their characteristics. Ignoring such complementarities results in misleading counterfactuals.

As shown in Proposition 1, merger simulations under linear pricing only require the identification of product-level market share functions. This implies that one may not need to point estimate the demand synergy parameters and the distribution of random coefficients in the GMM procedure to conduct such analyses. However, the estimation procedure in the current paper still assumes that the full model is identified. An interesting avenue for future research is to develop an adapted inference procedure for these counterfactuals that do not require the identification of the full model.

In practice, even though bundle-level market shares may not available, other bundle-level information may still be accessible. For example, household with membership card may receive a discount if she/he purchases a specific bundle of products. An extension of the current paper is to explore identification under endogenous and observed bundle-level discounts.

Similar to Allen and Rehbeck (2019a), the identification of product-level market share functions remains agnostic about whether a bundle is in the choice set, i.e. $\Gamma_{\mathbf{b}} \neq -\infty$. As shown in Theorem 3, one can identify whether $\Gamma_{\mathbf{b}} \neq -\infty$ in some models of demand for bundles. However, allowing for some $\Gamma_{\mathbf{b}}$ being $-\infty$ may introduce boundary problems in estimation and therefore complicates inference (see Andrews (2002)). In practice, an important question is how to select out (and test) those $\Gamma_{\mathbf{b}}$ that are potentially $-\infty$.

Finally, in the context of models of demand for single products, Reynaert and Verboven (2014) report remarkable efficiency gain by using optimal instruments. As mentioned in section 5, one may have non-trivial difficulties to construct the optimal instruments in the context of demand for bundles. An important question is whether and the extent to which a similar approach can be used to further improve the practical performance of the proposed methods.

References

Adams, W. J. and Yellen, J. L. (1976). Commodity bundling and the burden of monopoly. *The quarterly journal of economics*, pages 475–498.

Allen, R. and Rehbeck, J. (2019a). Identification with additively separable heterogeneity. *Econometrica*, 87(3):1021–1054.

Allen, R. and Rehbeck, J. (2019b). Latent complementarity in bundles models. *Available at SSRN 3257028*.

Amemiya, T. (1977). The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica (pre-1986)*, 45(4):955.

Andrews, D. W. (2017). Examples of l2-complete and boundedly-complete distributions. *Journal of econometrics*, 199(2):213–220.

Andrews, D. W. K. (2002). Generalized method of moments estimation when a parameter is on a boundary. *Journal of Business & Economic Statistics*, 20(4):530–544.

Armstrong, M. (2016a). Nonlinear pricing. *Annual Review of Economics*, 8:583–614.

Armstrong, M. and Vickers, J. (2010). Competitive non-linear pricing and bundling. *The Review of Economic Studies*, 77(1):30–60.

Armstrong, T. B. (2016b). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica*, 84(5):1961–1980.

Augereau, A., Greenstein, S., and Rysman, M. (2006). Coordination versus differentiation in a standards war: 56k modems. *The RAND Journal of Economics*, 37(4):887–909.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111.

Berry, S., Khwaja, A., Kumar, V., Musalem, A., Wilbur, K. C., Allenby, G., Anand, B., Chintagunta, P., Hanemann, W. M., Jeziorski, P., et al. (2014). Structural models of complementary choices. *Marketing Letters*, 25(3):245–256.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

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

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797.

Berto Villas-Boas, S. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *The Review of Economic Studies*, 74(2):625–652.

Bonnet, C. and Réquillart, V. (2013). Tax incidence with strategic firms in the soft drink market. *Journal of Public Economics*, 106:77–88.

Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). Database paper—the iri marketing data set. *Marketing science*, 27(4):745–748.

Canay, I. A., Santos, A., and Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81(6):2535–2559.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with pyblp. Technical report.

Cournot, A. (1838). *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan Co.

Crawford, G. S., Lee, R. S., Whinston, M. D., and Yurukoglu, A. (2018). The welfare effects of vertical integration in multichannel television markets. *Econometrica*, 86(3):891–954.

Crawford, G. S. and Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2):643–85.

D'Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, 27(3):460–471.

Dubé, J.-P. (2004). Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science*, 23(1):66–81.

Dubé, J.-P. H. (2018). Microeconometric models of consumer demand. Technical report, National Bureau of Economic Research.

Dubois, P., Griffith, R., and O'Connell, M. (forthcoming). How well targeted are soda taxes? *American Economic Review*.

Dunker, F., Hoderlein, S., and Kaido, H. (2017). Nonparametric identification of random coefficients in endogenous and heterogeneous aggregated demand models. Technical report, cemmap working paper.

Ershov, D., Laliberté, W. P., and Orr, S. (2018). Mergers in a model with complementarity. *Working Paper*, pages 1–50.

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review*, 103(5):1598–1628.

Fosgerau, M., Monardo, J., and de Palma, A. (2019). The inverse product differentiation logit model. *Working Paper*.

Fox, J. T. and Lazzati, N. (2017). A note on identification of discrete choice models for bundles and binary games. *Quantitative Economics*, 8(3):1021–1036.

Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181.

Freyberger, J. (2017). On completeness and consistency in nonparametric instrumental variable models. *Econometrica*, 85(5):1629–1644.

Gale, D. and Nikaido, H. (1965). The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159(2):81–93.

Gandhi, A. and Houde, J.-F. (2019). Measuring substitution patterns in differentiated products industries. Technical report, National Bureau of Economic Research.

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

Gillen, B. J., Montero, S., Moon, H. R., and Shum, M. (2019). Blp-2lasso for aggregate discrete choice models with rich covariates. *The Econometrics Journal*, 22(3):262–281.

Gordon, B. R. and Hartmann, W. R. (2013). Advertising effects in presidential elections. *Marketing Science*, 32(1):19–35.

Griffith, R., Nesheim, L., and O'Connell, M. (2018). Income effects and the welfare consequences of tax in differentiated product oligopoly. *Quantitative Economics*, 9(1):305–341.

Griffith, R., O'Connell, M., and Smith, K. (2019). Tax design in the alcohol market. *Journal of public economics*, 172:20–35.

Grzybowski, L. and Verboven, F. (2016). Substitution between fixed-line and mobile access: the role of complementarities. *Journal of Regulatory Economics*, 49(2):113–151.

Hendel, I. (1999). Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies*, 66(2):423–446.

Hu, Y. and Shiu, J.-L. (2018). Nonparametric identification using instrumental variables: sufficient conditions for completeness. *Econometric Theory*, 34(3):659–693.

Iaria, A. and Wang, A. (2019). Identification and estimation of demand for bundles. *Working Paper*, pages 1–92.

Ketz, P. (2019). On asymptotic size distortions in the random coefficients logit model. *Journal of Econometrics*.

Koijen, R. S. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, 127(4):1475–1515.

Kretschmer, T., Miravete, E. J., and Pernías, J. C. (2012). Competitive pressure and the adoption of complementary innovations. *American Economic Review*, 102(4):1540–70.

Kwak, K., Duvvuri, S. D., and Russell, G. J. (2015). An analysis of assortment choice in grocery retailing. *Journal of Retailing*, 91(1):19–33.

Lee, S., Kim, J., and Allenby, G. M. (2013). A direct utility model for asymmetric complements. *Marketing Science*, 32(3):454–470.

Liu, H., Chintagunta, P. K., and Zhu, T. (2010). Complementarities and the demand for home broadband internet services. *Marketing Science*, 29(4):701–720.

Manski, C. F. and Sherman, L. (1980). An empirical analysis of household choice among motor vehicles. *Transportation Research Part A: General*, 14(5-6):349–366.

Mattner, L. (1992). Completeness of location families, translated moments, and uniqueness of charges. *Probability Theory and Related Fields*, 92(2):137–149.

Mattner, L. (1993). Some incomplete but boundedly complete location families. *The Annals of Statistics*, 21(4):2158–2162.

Matutes, C. and Regibeau, P. (1992). Compatibility and bundling of complementary goods in a duopoly. *The Journal of Industrial Economics*, pages 37–54.

McAfee, R. P., McMillan, J., and Whinston, M. D. (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383.

Merlo, A. and Paula, Á. d. (2017). Identification and estimation of preference distributions when voters are ideological. *The Review of Economic Studies*, 84(3):1238–1263.

Milligan, K. and Rekkas, M. (2008). Campaign spending limits, incumbent spending, and election outcomes. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(4):1351–1374.

Mityagin, B. (2015). The zero set of a real analytic function. *arXiv preprint arXiv:1512.07276*.

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

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

Nevo, A., Rubinfeld, D. L., and McCabe, M. (2005). Academic journal pricing and the demand of libraries. *American Economic Review*, 95(2):447–452.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Ortega, J. M. (1968). The newton-kantorovich theorem. *The American Mathematical Monthly*, 75(6):658–660.

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

Pozzi, A. (2012). Shopping cost and brand exploration in online grocery. *American Economic Journal: Microeconomics*, 4(3):96–120.

Rekkas, M. (2007). The impact of campaign spending on votes in multiparty elections. *The Review of Economics and Statistics*, 89(3):573–585.

Reynaert, M. and Verboven, F. (2014). Improving the performance of random coefficients demand models: the role of optimal instruments. *Journal of Econometrics*, 179(1):83–98.

Samuelson, P. A. (1974). Complementarity: An essay on the 40th anniversary of the hicks-allen revolution in demand theory. *Journal of Economic literature*, 12(4):1255–1289.

Sher, I. and Kim, K. i. (2014). Identifying combinatorial valuations from aggregate demand. *Journal of Economic Theory*, 153:428–458.

Song, I. and Chintagunta, P. K. (2006). Measuring cross-category price effects with aggregate store data. *Management Science*, 52(10):1594–1609.

Thomassen, Ø., Smith, H., Seiler, S., and Schiraldi, P. (2017). Multi-category competition and market power: a model of supermarket pricing. *American Economic Review*, 107(8):2308–51.

Wang, A. (2020). Identifying the distribution of random coefficients in blp demand models using one single variation in product characteristics. *Available at SSRN 3569077*.

Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics*, 46(2):409–441.

Appendix

A Cross-Price Elasticities

For the cross-price elasticity between j and r:

$$\begin{aligned}
\frac{\partial s_{j.}}{\partial p_r} &= \int \sum_{\mathbf{b}:\mathbf{b}\ni j} \frac{\partial s_{i\mathbf{b}}}{\partial p_r} dF(\theta_i) \\
&= -\int \alpha_i \sum_{\mathbf{b}:\mathbf{b}\ni j} \frac{\partial s_{i\mathbf{b}}}{\partial \delta_{ir}} dF(\theta_i) \\
&= -\int \alpha_i \left[-\sum_{\mathbf{b}:\mathbf{b}\ni j, r\notin \mathbf{b}} s_{ir.} s_{i\mathbf{b}} + s_{ijr} - s_{ir.} s_{ijr} \right] dF(\theta_i) \\
&= \int \alpha_i \left[s_{ij.} s_{ir.} - s_{ijr} \right] dF(\theta_i).
\end{aligned}$$

B Proof of Proposition 1

Without loss of generality, I fix $x_{t\mathbf{J}} = x$ and drop it in the proof. Moreover, I also drop the notation t to simplify the exposition. Denote the ownership matrix in the factual by Ω and that after the merger by Ω_m.
Under Assumptions 1-2 and condition 4, due to Theorem 1, $\delta_{\mathbf{J}}$ are identified. Moreover, because (α, β) are identified, then $\xi_{\mathbf{J}}$ is identified.

First statement. For the price elasticity ϵ_{jr} at $p_{\mathbf{J}}$ between j and r, I obtain:

$$\epsilon_{jr} = \frac{p_j}{s_{j.}} \left[-\alpha \frac{\partial s_{j.}(\delta_{\mathbf{J}}; p_{\mathbf{J}}, \Gamma, F)}{\partial \delta_r} + \frac{\partial s_{j.}(\delta_{\mathbf{J}}; p_{\mathbf{J}}, \Gamma, F)}{\partial p_r} \right],$$

where $-\alpha\frac{\partial s_{j.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)}{\partial\delta_r}$ captures the variation of p_r that enters via the index δ_r, and $\frac{\partial s_{j.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)}{\partial p_r}$ captures the variation of p_r that enters via the unobserved deviation in u_{itr}. If α and $s_{j.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)$ are identified, then both $-\alpha\frac{\partial s_{j.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)}{\partial\delta_r}$ and $\frac{\partial s_{j.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)}{\partial p_r}$ are identified, and therefore, ϵ_{jr} is identified.

For the marginal costs $c_{\mathbf{J}}$, I first derive the first-order conditions (FOCs) of the Bertrand game in the factual:

$$\left[\Omega\odot\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}\right](p_{\mathbf{J}}-c_{\mathbf{J}})+s_{\mathbf{J}.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)=0.$$

First, note that $\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}=-\alpha\frac{\partial s_{\mathbf{J}.}}{\partial\delta_{\mathbf{J}}}+\frac{\partial s_{\mathbf{J}.}}{\partial p_{\mathbf{J}}}$ is identified. In addition, when $\alpha_i>0$ and condition 4 hold, as shown in the proof of Theorem 1 (see Appendix C), $\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}$ is negative definite and therefore $\Omega\odot\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}$ is also negative-definite and invertible. Then, $c_{\mathbf{J}}=p_{\mathbf{J}}+\left[\Omega\odot\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}\right]^{-1}s_{\mathbf{J}.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)$ is identified.

Given the uniqueness of the prices after the merger, it suffices to examine the FOCs of the Bertrand pricing game after the merger that uniquely determine the prices. In the case of mergers under linear pricing, the FOCs are:

$$\left[\Omega_m\odot\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}\right](p^m_{\mathbf{J}}-c_{\mathbf{J}})+s_{\mathbf{J}.}(\delta^m_{\mathbf{J}};p^m_{\mathbf{J}},\Gamma,F)=0, \tag{B.1}$$

where $p^m_{\mathbf{J}}$ are the prices after merger and $\delta^m_{\mathbf{J}}=-\alpha p^m_{\mathbf{J}}+\beta x+\xi_{\mathbf{J}}$ are the mean utilities of products after merger. Note that $\frac{\mathrm{d}s_{\mathbf{J}.}}{\mathrm{d}p_{\mathbf{J}}}$ is an identified function of $p^m_{\mathbf{J}}$. In addition, $\xi_{\mathbf{J}}$ and $c_{\mathbf{J}}$ are already identified. Then, $p^m_{\mathbf{J}}$ is uniquely determined by (B.1). Because the profit after merger is a function of $p^m_{\mathbf{J}}-c_{\mathbf{J}}$ and product-level market shares after merger, the profit change is also identified. Finally, denote the consumer surplus function by

$$V(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)=\mathbb{E}[u_{it\mathbf{b}}]=\mathbb{E}\left[\ln\left(1+\sum_{\mathbf{b}\in\mathbf{J}\cup\mathbf{C}_2}e^{\delta_{\mathbf{b}}(\Gamma_{\mathbf{b}})+\mu_{\mathbf{b}}(\theta_i;p_{\mathbf{J}})}\right)\right].$$

Note that $\frac{\partial V}{\partial\delta_{\mathbf{J}}}=s_{\mathbf{J}.}(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)$, for any $p_{\mathbf{J}}$. Then, for any $\delta_{\mathbf{J}}$ and $p_{\mathbf{J}}$, $\frac{\partial V}{\partial\delta_{\mathbf{J}}}$ and therefore $V(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)$ are identified. Given $(\alpha,\beta,\xi_{\mathbf{J}})$ and prices before and after the merger, the consumer surplus change, $\Delta V=V(\delta^m_{\mathbf{J}};p^m_{\mathbf{J}},\Gamma,F)-V(\delta_{\mathbf{J}};p_{\mathbf{J}},\Gamma,F)$, is identified.

Second statement. First, because $c_{t\mathbf{b}}=\sum_{j\in\mathbf{b}}c_{tj}$, then bundle-level marginal costs are also identified. When there is nonlinear pricing after the merger, bundle-level market share functions will further depend on the discounts or surcharges set by firms, which is written as $s_{\mathbf{b}}(\delta_{\mathbf{C}_1};p_{\mathbf{C}_1},F)$, where $p_{\mathbf{C}_1}=(p_{\mathbf{J}},p_{\mathbf{C}_2})$.

Suppose (α, β, Γ) are identified. For the first case in which F is further identified, then $s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; p_{\mathbf{C}_1}, F)$ is directly identified as a function of $(\delta_{\mathbf{C}_1}, p_{\mathbf{C}_1})$. For the second case, note that if $\alpha_i = \alpha$, then prices enter market share functions only via the mean utilities of products or bundles. As a consequence, bundle-level market share functions $s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; p_{\mathbf{C}_1}, F)$ do not directly depend on $p_{\mathbf{C}_1}$ and they can be re-written as: $s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; p_{\mathbf{C}_1}, F) = s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; F)$. Similarly, in the factual scenario, $s_{\mathbf{b}}(\delta_t; p_{t\mathbf{J}}, F)$ can be re-written as $s_{\mathbf{b}}(\delta_t; p_{t\mathbf{J}}, F) = s_{\mathbf{b}}(\delta_t; F)$. If identifying $s_{\mathbf{b}}(\delta_t; F)$ in the factual implies the identification of $s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; F)$ in the counterfactual. In either case, one can then apply similar arguments to those in the proof of the first statement to mergers under nonlinear pricing.

C Proof of Theorem 1

We need the following regularity condition:

Assumption 4. $\frac{\partial s(\delta_t; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \theta_{it})}{\partial \delta_t}$ *is symmetric and positive definite.*

In fact, when Φ is continuous, the inclusive value function of individual i in market t, $\int \max\{u_{it\mathbf{b}}\}_{\mathbf{b}\in\mathbf{C}} d\Phi(\varepsilon_{it})$, is already convex function of δ_t.[49] As a result, the Hessian matrix of the inclusive value function, $\frac{\partial s(\delta_t; x^{(2)}_{t\mathbf{J}}, p_{t\mathbf{J}}, \theta_{it})}{\partial \delta_t}$, is symmetric and semi-positive definite. This assumption strengthens the Hessian matrix to be positive-definite. This assumption is generically true and holds for often used distributions Φ (e.g. Gumbel, Gaussian).
I drop the notation t to simplify the exposition. Denote by $\mathbf{W}$ a matrix of dimension $J \times C_1$ and the jth row is $\mathbf{w}_j$. Note that the first $J \times J$ component is the identify matrix of size $J \times J$. Therefore, $\mathbf{W}$ is of full row rank. We can re-write (4) as:

$$\begin{aligned} s_{\mathbf{J}\cdot} &= s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x^{(2)}_{\mathbf{J}}, p_{\mathbf{J}}, \Gamma, F) \\ &= \mathbf{W}\ s(\mathbf{W}^{\mathrm{T}}\delta_{\mathbf{J}} + (0, ..., 0, \Gamma)^{\mathrm{T}}; x^{(2)}_{\mathbf{J}}, p_{\mathbf{J}}, F) \end{aligned} \tag{C.1}$$

Then,

$$\frac{\partial s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x^{(2)}_{\mathbf{J}}, p_{\mathbf{J}}, \Gamma, F)}{\partial \delta_{\mathbf{J}}} = \mathbf{W}\ \frac{\partial s(\mathbf{W}^{\mathrm{T}}\delta_{\mathbf{J}} + (0, ..., 0, \Gamma)^{\mathrm{T}}; x^{(2)}_{\mathbf{J}}, p_{\mathbf{J}}, F)}{\partial \delta}\ \mathbf{W}^{\mathrm{T}}. \tag{C.2}$$

Note that
$$\frac{\partial s(\delta; x^{(2)}_{\mathbf{J}}, p_{\mathbf{J}}, F)}{\partial \delta} = \int \frac{s(\delta; x^{(2)}_{\mathbf{J}}, p_{\mathbf{J}}, \theta_i)}{\partial \delta} dF(\theta_i).$$

[49] Or equivalently, the expenditure function of individual i in market t is concave in δ_t, when the price coefficient α_i is negative.

Because of Assumption 4, $\frac{s(\delta;x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\theta_i)}{\partial\delta}$ is symmetric and positive definite for each θ_i. Then, $\frac{\partial s(\delta;x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},F)}{\partial\delta}$ is also symmetric and positive definite. Moreover, $\mathbf{W}$ is of full row rank. Consequently, $\frac{\partial s_{t\mathbf{J}.}(\delta_{\mathbf{J}};x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma,F)}{\partial\delta_{\mathbf{J}}}$ is still symmetric and positive definite, and therefore positive quasi-definite. Then, given $(x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma,F)$, for $\delta_{\mathbf{J}}\in\mathbb{R}^J$, according to Theorem 6 of Gale and Nikaido (1965), $s_{\mathbf{J}.}(\delta_{\mathbf{J}};x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma,F)$ defines a bijection from $\mathbb{R}^J$ to $s_{\mathbf{J}.}(\mathbb{R}^J;x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma,F)$.

Similarly, for any $(x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma',F')$, $s_{\mathbf{J}.}(\delta_{\mathbf{J}};x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma',F')$ defines a bijection from $\mathbb{R}^J$ to $s_{\mathbf{J}.}(\mathbb{R}^J;x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma',F')$. If the true $\mathcal{s}_{\mathbf{J}.}\in s_{\mathbf{J}.}(\mathbb{R}^J;x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma',F')$, then there is a unique $\delta'_{\mathbf{J}}$ such that $\mathcal{s}_{\mathbf{J}.}=s_{\mathbf{J}.}(\delta'_{\mathbf{J}};x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma',F')$. If $\mathcal{s}_{\mathbf{J}.}\notin s_{\mathbf{J}.}(\mathbb{R}^J;x^{(2)}_{\mathbf{J}},p_{\mathbf{J}},\Gamma',F')$, then there is no such $\delta'_{\mathbf{J}}$. The proof is completed.

D Proof of Theorem 2

In what follows, I fix $x^{(2)}_{t\mathbf{J}}=x^{(2)}$ and prove the conclusion for each $x^{(2)}\in\mathbf{D}^{(2)}_x$. To start with, I plug the definition of $\xi_{\mathbf{J}}(\cdot)$ into (8):

$$\mathbb{E}[s^{-1}_{\mathbf{J}.}(\mathcal{s}_{t\mathbf{J}.};x^{(2)},p_{t\mathbf{J}},\Gamma,F)+\alpha p_{t\mathbf{J}}-\beta x_{t\mathbf{J}}|z_{t\mathbf{J}}=z,x_{t\mathbf{J}}=x]=0, \tag{D.1}$$

where $z\in\mathbf{D}_z$, $x\in\mathbf{D}_x=\mathbf{D}^{(1)}_x\times\mathbf{D}^{(2)}_x$, and $\mathbf{D}_z$ denote the support of $z_{t\mathbf{J}}$. I also assume the following regularity condition:

Condition 1. *For any (Γ',F') and any $z\in\mathbf{D}_z$, there exists $M_z>0$, such that*

$$\mathbb{E}\left[\left|s^{-1}_{\mathbf{J}.}(\mathcal{s}_{t\mathbf{J}.};x^{(2)},p_{t\mathbf{J}},\Gamma,F)\right|\,\Big|z\right],\mathbb{E}\left[|p_{t\mathbf{J}}|\,\Big|z\right]\le M_z.$$

Sketch of the proof. The proof is proceeded in three steps. In the first step, I prove that under conditions 1-3 in Theorem 2, the identification by moment restrictions (8) is equivalent to uniquely solving a convolution equation. This convolution equation is generated by the distribution of demand and shocks, and the translation in the convolution equation is defined by $z_{t\mathbf{J}}\in\mathbb{R}^J$. In the second step, using condition 4 in Theorem 2 and that $\mathbf{D}^{(1)}_x$ and $\mathbf{D}^{(2)}_x$ are both open, I prove that the property that the zero function is the unique solution to the convolution equation is sufficient for the identification of α and $s_{\mathbf{J}.}(\delta_{\mathbf{J}};x^{(2)},\Gamma,F)$ as functions of $(\delta_{\mathbf{J}},x^{(2)})\in\mathbb{R}^J\times\mathbf{D}^{(2)}_x$. In the final step, by leveraging the completeness of location families in Mattner (1992), I demonstrate that when $(\xi_{t\mathbf{J}},w_{t\mathbf{J}})$ is Gaussian distributed (or their joint distribution satisfies some "fat-tail" conditions), the property that the zero function is the unique solution to the convolution equation will hold; under regularity condition 1 of Appendix D, by leveraging the polynomial completeness in D'Haultfoeuille (2011), I prove that the same property will hold when the DGP is a model of demand for multiple products across categories.

D.1 Moment Restrictions (D.1) and Convolution Equation

When $\alpha_i = \alpha$, $s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x^{(2)}, p_{t\mathbf{J}}, \Gamma, F)$ depends on $p_{t\mathbf{J}}$ only via the index $\delta_{t\mathbf{J}}$. Then, $s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x^{(2)}, p_{t\mathbf{J}}, \Gamma, F)$ can be written as $s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; x^{(2)}, \Gamma, F)$. The following theorem transforms (D.1) to a convolution equation:

Theorem D.1. *Suppose that Assumptions 1-3 and regularity condition 1 hold. Moreover, the following conditions hold:*

1. $z_{t\mathbf{J}}$ *is independent of* $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$.
2. $\alpha_i = \alpha \neq 0$
3. *Given* $x^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x_{t\mathbf{J}} + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}; x^{(2)})$ *is a continuous function of* $(\beta x_{t\mathbf{J}} + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$.

Then, for any $z \in \mathbf{D}_z$ *and* $x \in \mathbf{D}_x$, $(\alpha', \beta', \Gamma', F')$ *satisfies moment conditions* (D.1) *if and only if the following convolution equation*

$$\int G(t; \alpha', \beta, \Gamma', F')\Lambda_G(t - z; f_{\xi,w})dt = 0, \tag{D.2}$$

holds, where

$$\begin{aligned} G(t; \alpha', \beta, \Gamma', F') &= s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, t; x^{(2)}); x^{(2)}, \Gamma, F); x^{(2)}, \Gamma', F') \\ &\quad + \alpha' p_{\mathbf{J}}(0, t; x^{(2)}) + \left(\frac{\alpha'}{\alpha}\beta - \beta\right) x, \\ \Lambda_G(\lambda; f_{\xi,w}) &= \int \alpha f_{\xi,w}(\alpha(w - \lambda) - \beta x, w)dw, \end{aligned}$$

and $f_{\xi,w}$ *is the density function of* $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$.

Proof. Since $x^{(2)}$ is fixed, I drop this notation in this proof and also the notation of $p_{\mathbf{J}}(\cdot)$ and $s_{\mathbf{J}\cdot}(\cdot; \Gamma', F')$. To start with, I prove the following Lemma:

Lemma 1. *Suppose that* $\alpha_i = \alpha$ *and* $p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$ *is a function of* $(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$. *Then, for any* $\Delta \in \mathbb{R}^J$,

$$p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}} + \alpha\Delta, c_{t\mathbf{J}} + \Delta) = p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \Delta.$$

Proof. Denote by Ω the factual ownership matrix. Then, I can derive the FOCs of the simultaneous Bertrand pricing game:

$$-\alpha \left[\Omega \odot \frac{\partial s_{\mathbf{J}\cdot}}{\partial \delta_{t\mathbf{J}}}\right](p_{t\mathbf{J}} - c_{t\mathbf{J}}) + s_{\mathbf{J}\cdot}(\delta_{t\mathbf{J}}; \Gamma, F) = 0, \tag{D.3}$$

where $\delta_{t\mathbf{J}} = -\alpha p_{t\mathbf{J}} + \beta x + \xi_{t\mathbf{J}}$. Suppose that $c_{t\mathbf{J}}$ increases by Δ and $\beta x + \xi_{t\mathbf{J}}$ increases by $\alpha\Delta$. Then, the FOCs (D.3) with $p'_{t\mathbf{J}} = p_{\mathbf{J}}(\xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \Delta$, $c'_{t\mathbf{J}} = c_{t\mathbf{J}} + \Delta$ and $\delta_{t\mathbf{J}}$ still hold because $\delta_{t\mathbf{J}}$ and $p_{t\mathbf{J}} - c_{t\mathbf{J}}$ remain unchanged. Due to the uniqueness of $p_{t\mathbf{J}}$ as function of $(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$, I obtain that $p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}} + \alpha\Delta, c_{t\mathbf{J}} + \Delta) = p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}) + \Delta$. □

First, I prove the sufficiency part of Theorem D.1. For any $\Delta \in \mathbb{R}^J$, by using Lemma 1, I obtain:

$$\begin{aligned}
\mathbb{E}[p_{t\mathbf{J}} | z_{t\mathbf{J}} = z, x_{t\mathbf{J}} = x] &= \int p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}}, z + w_{t\mathbf{J}}) f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \\
&= \int \left[p_{\mathbf{J}}(0, z + w_{t\mathbf{J}} - \frac{\beta x + \xi_{t\mathbf{J}}}{\alpha}) + \frac{\beta x + \xi_{t\mathbf{J}}}{\alpha} \right] f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \\
&= \int p_{\mathbf{J}}(0, z + w_{t\mathbf{J}} - \frac{\beta x + \xi_{t\mathbf{J}}}{\alpha}) f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) + \frac{\beta x}{\alpha}
\end{aligned} \tag{D.4}$$

Similarly, for $(\alpha', \beta', \Gamma', F')$ satisfying (D.1), I compute

$$\begin{aligned}
&\mathbb{E}[s_{\mathbf{J}\cdot}^{-1}(\mathscr{s}_{t\mathbf{J}\cdot}; \Gamma', F') | z_{t\mathbf{J}} = z] \\
&= \int s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(\beta x + \xi_{t\mathbf{J}} - \alpha p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}}, z_{t\mathbf{J}} + w_{t\mathbf{J}}); \Gamma, F); \Gamma', F') f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) \\
&= \int s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, z + w_{t\mathbf{J}} - \frac{\beta x + \xi_{t\mathbf{J}}}{\alpha}); \Gamma, F); \Gamma', F') f_{\xi,w}(\xi_{t\mathbf{J}}, w_{t\mathbf{J}}) d(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})
\end{aligned} \tag{D.5}$$

I now plug (D.4) and (D.5) in (D.1) evaluated at $(\alpha', \beta', \Gamma', F')$, and transform $(\xi_{t\mathbf{J}}, w_{t\mathbf{J}})$ to $(z + w_{t\mathbf{J}} - \frac{\beta x + \xi_{t\mathbf{J}}}{\alpha}, w_{t\mathbf{J}})$:

$$\begin{aligned}
&\mathbb{E}[s_{\mathbf{J}\cdot}^{-1}(\mathscr{s}_{t\mathbf{J}\cdot}; \Gamma', F') + \alpha' p_{t\mathbf{J}} - \beta' x | z_{t\mathbf{J}} = z] \\
&= \int [s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, t); \Gamma, F); \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t) + \left(\frac{\alpha'}{\alpha}\beta - \beta' \right) x] \alpha f_{\xi,w}(\alpha(z + w_{t\mathbf{J}} - t) - \beta x, w_{t\mathbf{J}}) d(t, w_{t\mathbf{J}}) \\
&= \int [s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, t); \Gamma, F); \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t) + \left(\frac{\alpha'}{\alpha}\beta - \beta' \right) x] \Lambda_G(t - z; f_{\xi,w}) dt.
\end{aligned}$$

The proof of sufficiency is completed. To obtain the necessity part, one can reverse the argument of the sufficiency. □

D.2 Unique Solution for Convolution Equation (D.2) and Identification of Product-Level Market Share Functions

Denote the identification set derived from moment conditions (D.1) by

$$\Theta = \{(\alpha', \beta', \Gamma', F') : \text{(D.1) holds at } (\alpha', \beta', \Gamma', F') \text{ for any } z \in \mathbf{D}_z, x \in \mathbf{D}_x\},$$

and that derived from (D.2) by

$$\Theta_G = \{(\alpha', \beta', \Gamma', F') : \text{(D.2) holds for any } z \in \mathbf{D}_z, x \in \mathbf{D}_x\}.$$

Theorem D.1 establishes $\Theta = \Theta_G$. Define $\Theta_G^0 = \{(\alpha', \beta', \Gamma', F') : G(\cdot; \alpha', \beta', \Gamma', F') = 0\}$, the set of parameters that deliver $G(\cdot) = 0$. Note that $\Theta_G^0 \subset \Theta_G$ and the true parameters $(\alpha, \beta, \Gamma, F) \in \Theta_G^0 \subset \Theta_G = \Theta$. Then, a necessary condition for the identification of $(\alpha, \beta, \Gamma, F)$ by moment conditions (D.1), i.e. $\Theta = \{(\alpha, \beta, \Gamma, F)\}$, is $\Theta_G^0 = \Theta_G$, i.e. $G = 0$ is the unique solution of convolution equation (D.2). This is the completeness of the location families generated by $\Lambda_G(\cdot; f_{\xi,w})$. The next theorem characterises the implications of this completeness for identification:

Theorem D.2. *Suppose that conditions of Theorem D.1 hold.*

1. *If $\Theta = \{(\alpha, \beta, \Gamma, F)\}$, then $\Theta_G^0 = \Theta_G$.*
2. *If $\Theta_G^0 = \Theta_G$, then (α, β) are identified and $s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x^{(2)}, \Gamma, F)$ are identified as functions of $(\delta_{\mathbf{J}}, x^{(2)}) \in \mathbb{R}^J \times \mathbf{D}_x^{(2)}$.*

Remark 4. *The first statement of Theorem D.2 shows that the completeness of the location families ($\Theta_G^0 = \Theta_G$) is necessary for the identification of the full model by moment conditions* (8)*. The second statement shows that the completeness condition is also sufficient for the identification of (α, β) and $s_{\mathbf{J}\cdot}(\cdot; x^{(2)}, \Gamma, F)$.*

Proof. We have proven the first statement in the previous paragraph. We prove the second statement for any fixed $x^{(2)}$. Note that if $G(t; \alpha', \beta', \Gamma', F') = 0$ for any $t \in \mathbb{R}^J$, then we have

$$s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, t); \Gamma, F) = s_{\mathbf{J}\cdot}(-\alpha' p_{\mathbf{J}}(0, t) + v; \Gamma', F'), \tag{D.6}$$

for any $t \in \mathbb{R}^J$, where $v = \left(\frac{\alpha'}{\alpha}\beta - \beta'\right) x$.

We first prove that $\mathbf{D}_p = \{p' \in \mathbb{R}^J : p' = p_{\mathbf{J}}(0, t), t \in \mathbb{R}^J\}$ is an open set in $\mathbb{R}^J$. As shown in the proof of Proposition 1 (Appendix B), marginal costs $c_{t\mathbf{J}}$ are identifiable: for any $p'_{t\mathbf{J}}$, there exists a unique $c'_{t\mathbf{J}}$ such that the FOCs of the Bertrand pricing game hold. Moreover, this mapping from $p'_{t\mathbf{J}}$ to $c'_{t\mathbf{J}}$ is C^1. Because $p_{\mathbf{J}}(0, c'_{t\mathbf{J}})$ is a continuous function, then $p'_{t\mathbf{J}} = p_{\mathbf{J}}(0, c'_{t\mathbf{J}})$ defines a continuous bijection between prices and marginal costs. Consequently, $\mathbf{D}_p$ is an open set in $\mathbb{R}^J$ and (D.6) holds in $\mathbf{D}_p$.

According to Iaria and Wang (2019) (Theorem Real Analytic Property), given any (Γ'', F''), $s_{\mathbf{b}}(\delta'_{t\mathbf{J}}; \Gamma'', F'')$ is real analytic with respect to $\delta'_{t\mathbf{J}}$. Then, $s_{\mathbf{J}\cdot}(\delta'_{t\mathbf{J}}; \Gamma'', F'')$ is real analytic with respect to $\delta'_{t\mathbf{J}}$. Consequently, $s_{\mathbf{J}\cdot}(-\alpha p'_{t\mathbf{J}}; \Gamma, F)$ and $s_{\mathbf{J}\cdot}(-\alpha' p'_{t\mathbf{J}} + v; \Gamma', F')$ are both real analytic with respect to $p'_{t\mathbf{J}}$. Because these two real analytic functions coincide on open set $\mathbf{D}_p \subset \mathbb{R}^J$, then they coincide for all $p_{t\mathbf{J}} \in \mathbb{R}^J$: (D.6) holds for any $p_{t\mathbf{J}} \in \mathbb{R}^J$.

We now prove $\alpha = \alpha'$. To do this, we first let the p_{tj}'s, $j \neq 1$, tend to infinity. Then, we obtain a choice set that only include product 1 and the outside option 0. As a consequence, Γ (Γ') do not enter the left-hand (right-hand) side of (D.6). Then,

$$\begin{aligned} s_1(-\alpha p_1; F) &= \int \frac{\exp\{-\alpha p_1 + \mu_1(\theta_i)\}}{1+\exp\{-\alpha p_1 + \mu_1(\theta_i)\}} dF(\theta_i) \\ &= s_1(-\alpha' p_1 + v; F') \\ &= \int \frac{\exp\{-\alpha' p_1 + v + \mu_1(\theta_i)\}}{1+\exp\{-\alpha' p_1 + v + \mu_1(\theta_i)\}} dF'(\theta_i). \end{aligned} \tag{D.7}$$

We prove $\alpha = \alpha'$ by contradiction. Without loss of generality, suppose that $0 < \alpha' < \alpha$. Denote the compact supports of F and F' by $\mathbf{D}_F$ and $\mathbf{D}_{F'}$, respectively. Then, for sufficiently large $p_1 > 0$, the set $\{-\alpha p_1 + \mu_1(\theta_i) : \theta_i \in \mathbf{D}_F\}$ will be completely on the left of $\{-\alpha' p_1 + v + \mu_1(\theta_i) :, \theta_i \in \mathbf{D}_{F'}\}$.[50] Since the function $\exp\{x\}/(1+\exp\{x\})$ is a strictly increasing function, we obtain that $s_1(-\alpha p_1; F) < s_1(-\alpha' p_1 + v; F')$, which contradicts (D.7). $\alpha = \alpha'$ is proved.

Plugging $\alpha' = \alpha$ in (D.6), we obtain that for any $p \in \mathbb{R}^J$:

$$s_{\mathbf{J}\cdot}(-\alpha p; \Gamma, F) = s_{\mathbf{J}\cdot}(-\alpha p + v; \Gamma', F').$$

where v on the right-hand side is equal to $\left(\frac{\alpha'}{\alpha}\beta - \beta'\right) x = (\beta - \beta')x$. Because $x^{(1)}$ varies in open set $\mathbf{D}_x^{(1)}$, then we obtain $\beta^{(1)} = \beta'^{(1)}$. We now prove $\beta^{(2)} = \beta'^{(2)}$. Note that given $\beta^{(1)} = \beta'^{(1)}$ and $\alpha = \alpha'$, we have:

$$s_{\mathbf{J}\cdot}(-\alpha p + \beta^{(2)} x^{(2)}; \Gamma, F) = s_{\mathbf{J}\cdot}(-\alpha p + \beta'^{(2)} x^{(2)}; \Gamma', F').$$

We re-use the technique used to prove $\alpha = \alpha'$:

$$\int \frac{\exp\{-\alpha p_1 + (\beta^{(2)} + \Delta\beta_i^{(2)})x^{(2)} + \eta_{i1}\}}{1+\exp\{-\alpha p_1 + (\beta^{(2)} + \Delta\beta_i^{(2)})x^{(2)} + \eta_{i1}\}} dF(\theta_i) = \int \frac{\exp\{-\alpha p_1 + (\beta^{(2)} + \Delta\beta_i^{(2)})x^{(2)} + \eta_{i1}\}}{1+\exp\{-\alpha p_1 + (\beta^{(2)} + \Delta\beta_i^{(2)})x^{(2)} + \eta_{i1}\}} dF'(\theta_i). \tag{D.8}$$

Then, according to Theorem 1 of Wang (2020), the distribution of $(\beta^{(2)} + \Delta\beta_i^{(2)})x^{(2)} + \eta_{i1}$ conditional on $x^{(2)}$ is identified. Then, its mean (conditional on $x^{(2)}$) is also identified: $\beta^{(2)} x^{(2)} + \mathbb{E}_F[\eta_{i1}] = \beta'^{(2)} x^{(2)} + \mathbb{E}_{F'}[\eta_{i1}]$. As a result, for any $x^{(2)}, x'^{(2)} \in \mathbf{D}_x^{(2)}$, we obtain $(\beta^{(2)} - \beta'^{(2)})(x^{(2)} - x'^{(2)}) = 0$. Since $\mathbf{D}_x^{(2)}$ is open, we obtain $\beta^{(2)} = \beta'^{(2)}$ and $v = 0$. The proof holds for any $x^{(2)} \in \mathbf{D}_x^{(2)}$. Finally, we proved $s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; \Gamma, F) =$

[50] When F and F' have unbounded support, the two sets are not disjoint. However, if the tails of F and F' are sufficiently thin, then the probability weights are mostly concentrated around $-\alpha p_1 + \mu_F$ and $-\alpha' p_1 + \mu_{F'}$, where μ_F and $\mu_{F'}$ are the mean of F and F', respectively. Since $-\alpha p_1 + \mu_F$ and $-\alpha' p_1 + \mu_{F'}$ are still sufficiently distant, the argument here will still go through.

$s_{\mathbf{J}.}(\delta_{\mathbf{J}};\Gamma',F')$ for $(\delta_{\mathbf{J}},x^{(2)})\in\mathbb{R}^J\times\mathbf{D}_x^{(2)}$. The proof is completed. □

D.3 Sufficient Conditions for the Completeness of Location Families

In general, depending on the regularity of $G(\cdot)$ (bounded, polynomially bounded, integrable with respect to $\Lambda_G(\cdot)$, etc.), the completeness of location families can be achieved with different sufficient conditions on $\Lambda_G(\cdot)$ (and hence on $f_{\xi,w}$).[51] The next theorem establishes two sets of sufficient conditions for the completeness of location families ($\Theta_G^0=\Theta_G$):

Theorem D.3. *Suppose that* (D.2) *holds for $z\in\mathbb{R}^J$.*

- *If $f_{\xi,w}$ is Gaussian, then $\Theta_G^0=\Theta_G$.*
- *If the DGP is a model of demand for multiple products across categories, then under regularity conditions 2, $\Theta_G^0=\Theta_G$.*

Proof. Note that the location families are generated by $\Lambda_G(\cdot;f_{\xi,w})$, which is the density function of a translation of demand and supply shocks in model (4). When $f_{\xi,w}$ is Gaussian, $\Lambda_G(\lambda;f_{\xi,w})$ is also Gaussian. Then, the first statement follows directly from Theorem 2.4 of Mattner (1993).
For the second statement, I leverage Theorem 2.1 of D'Haultfoeuille (2011). To do so, I require the following regularity conditions:

Condition 2.

(i). $(\xi_{t\mathbf{J}},w_{t\mathbf{J}})$ *are continuous random variables with finite moments.*

(ii). The characteristics function of Λ_G is infinitely often differentiable in $\mathbb{R}^J$ except for some finite set. Moreover, the characteristics function of Λ_G does not vanish on $\mathbb{R}^J$.

(iii). There exists B and l, such that $|p_{t\mathbf{J}}(0,c_{t\mathbf{J}})-p_{t\mathbf{J}}(0,0)|\leq B|c_{t\mathbf{J}}|^l$, where $|.|$ refers to Euclidean norm.

Condition 2(i) implies Assumption A3 of D'Haultfoeuille (2011). Condition 2(ii) implies his Assumption A4. The differentiability requirement and the zero-freeness requirement are satisfied by many commonly used distributions. Condition 2(iii) restricts pricing behaviours to be controlled by a polynomial of marginal costs and is satisfied at least by mixed logit models of demand for single products. Moreover, together with Condition 2(i), it implies Assumption A5 of D'Haultfoeuille (2011).

[51] For different concepts of completeness, see Mattner (1992, 1993), D'Haultfoeuille (2011), and Andrews (2017).

First, I re-write G as a function of $p_{t\mathbf{J}}(0,t)$:

$$G = G(p_{t\mathbf{J}}) = s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{t\mathbf{J}}; x^{(2)}, \Gamma, F); x^{(2)}, \Gamma', F') + \alpha' p_{t\mathbf{J}} + \left(\frac{\alpha'}{\alpha}\beta - \beta'\right) x.$$

To apply statement (ii) of Theorem 2.1 in D'Haultfoeuille (2011), it is enough to prove that G can be polynomially controlled by $p_{t\mathbf{J}}$:

Lemma 2. *There exists $A, M > 0$, such that $|G(p_{t\mathbf{J}})| \leq A|p_{t\mathbf{J}}| + M$, for any $p_{t\mathbf{J}} \in \mathbb{R}^J$.*

Combining this lemma with Conditions 2(i)-(iii), I can apply the P-completeness result in Theorem 2.1 of D'Haultfoeuille (2011): if G satisfies convolution equation (D.2) for any $z \in \mathbb{R}^J$, then $G \equiv 0$. In the remaining part, I prove Lemma 2.

Without loss of generality, normalize the support of F and F' to $[0,1]^R$, where R is the dimension of random coefficients. Note that it is sufficient to prove

$$s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F')| \leq A'|p_{t\mathbf{J}}| + M'$$

for some constant A' and M'. First, consider a model of demand for single products. For any $\delta'_{\mathbf{J}} \in \mathbb{R}^J$, denote $\mathcal{s}_{\mathbf{J}} = s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; F) = s_{\mathbf{J}\cdot}(\delta'_{\mathbf{J}}; F')$. Then, we have for any $j \in \mathbf{J}$:

$$\begin{aligned} \ln \mathcal{s}_j - \ln \mathcal{s}_0 &= \delta_j + \ln \frac{\int \frac{e^{\mu_{tj}(\theta_i)}}{1+\sum_{j\in\mathbf{J}} e^{\delta_j + \mu_{tj}(\theta_i)}} dF(\theta_i)}{\int \frac{1}{1+\sum_{j\in\mathbf{J}} e^{\delta_j + \mu_{tj}(\theta_i)}} dF(\theta_i)} \\ &= \delta_j + \mu_{tj}(\tilde{\theta}), \end{aligned}$$

where $\tilde{\theta}$ is some value in $[0,1]^R$. We apply the same arguments to F' and obtain:

$$\ln \mathcal{s}_j - \ln \mathcal{s}_0 = \delta'_j + \mu_{tj}(\tilde{\theta}'),$$

where $\tilde{\theta}'$ is some value in $[0,1]^R$. Then, we have $\delta_j - \delta'_j = \mu_{tj}(\tilde{\theta}') - \mu_{tj}(\tilde{\theta})$. Because both $\tilde{\theta}$ and $\tilde{\theta}'$ are bounded by 1, then $\mu_{tj}(\tilde{\theta})$ and $\mu_{tj}(\tilde{\theta}')$ are also bounded. As a result, we obtain that there exists a constant M_j that does not depend on $\delta_{\mathbf{J}}$, such that $|\delta_j - \delta'_j| \leq M_j$. Consequently, $|\delta - \delta'| \leq M' = \sqrt{\sum_{j=1}^J M_j^2}$, or equivalently, $|s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; \Gamma, F); \Gamma', F') - \delta_{\mathbf{J}}| \leq M'$ for any $\delta_{\mathbf{J}} \in \mathbb{R}^J$. Plug $\delta_{\mathbf{J}} = -\alpha p_{t\mathbf{J}}$ into this inequality, we obtain $|s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F') + \alpha p_{t\mathbf{J}}| \leq M'$ and therefore $|s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{t\mathbf{J}}; \Gamma, F); \Gamma', F')| \leq \alpha |p_{t\mathbf{J}}| + M'$.

For models of demand for multiple products across K categories, for any $\delta'_{\mathbf{J}} \in \mathbb{R}^J$, denote $\mathcal{s}_{\mathbf{J}} = s_{\mathbf{J}\cdot}(\delta_{\mathbf{J}}; x^{(2)}, \Gamma, F) = s_{\mathbf{J}\cdot}(\delta'_{\mathbf{J}}; x^{(2)}, \Gamma', F')$. Take product category $\mathbf{J}_1$ and define $\tilde{\jmath}_0 = 1 - \sum_{j\in\mathbf{J}_1} \jmath_{j\cdot}$. Note that

$$\tilde{\jmath}_0 = \int \frac{\sum_{\mathbf{b}=((j_k)_{j_k\in\mathbf{J}_k,k=2,\dots,K})} e^{\sum_{k=2}^{K}\delta_{j_k}+\Gamma_{\mathbf{b}}+\sum_{k=2}^{K}\mu_{tj_k}(\theta_i)}}{\sum_{\mathbf{b}=(j_k)_{j_k\in\mathbf{J}_k,k=1,\dots,K}} e^{\sum_{k=1}^{K}\delta_{j_k}+\Gamma_{\mathbf{b}}+\mu_{t\mathbf{b}}(\theta_i)}} dF(\theta_i)$$

and for $j \in \mathbf{J}_1$,

$$\jmath_{j\cdot} = \int \frac{\sum_{\mathbf{b}=((j_k)_{j_k\in\mathbf{J}_k,k=2,\dots,K})} e^{\delta_j+\mu_{tj}(\theta_i)} e^{\sum_{k=2}^{K}\delta_{j_k}+\Gamma_{\mathbf{b}\cup\{j\}}+\sum_{k=2}^{K}\mu_{tj_k}(\theta_i)}}{\sum_{\mathbf{b}=(j_k)_{j_k\in\mathbf{J}_k,k=1,\dots,K}} e^{\sum_{k=1}^{K}\delta_{j_k}+\Gamma_{\mathbf{b}}+\mu_{t\mathbf{b}}(\theta_i)}} dF(\theta_i).$$

Then, similar to demand models of single products, we obtain:

$$\ln \jmath_{j\cdot} - \ln \tilde{\jmath}_0 = \delta_j + \mu_{tj}(\tilde{\theta}) + \Gamma_{\tilde{\mathbf{b}}\cup\{j\}} - \Gamma_{\tilde{\mathbf{b}}},$$

where $\tilde{\theta}$ is some value in $[0,1]^R$ and $\tilde{\mathbf{b}}$ is some bundle without $j \in \mathbf{J}_1$. We apply the same arguments to (Γ', F') and obtain:

$$\ln \jmath_{j\cdot} - \ln \jmath_0 = \delta'_j + \mu_{tj}(\tilde{\theta}') + \Gamma_{\tilde{\mathbf{b}}'\cup\{j\}} - \Gamma_{\tilde{\mathbf{b}}'},$$

where $\tilde{\theta}'$ is some value in $[0,1]^R$ and $\tilde{\mathbf{b}}'$ is some bundle without $j \in \mathbf{J}_1$. Then, similar arguments in the model of demand for single product apply and $|\delta_j - \delta'_j|$ is bounded by some constant that only depends on the support of $\tilde{\theta}$ and $\tilde{\theta}'$ and the bounds of Γ. The proof of the second statement is completed. □

Combining Theorems D.1-D.3, we obtain Theorem 2.

E Proof of Theorem 3

The proof is proceeded in two steps. In the first step, I prove $\Gamma_{(j,r)}$ is identified for $j, r \in \mathbf{J}$, $j \neq r$. In the second step, I prove that the $s_{\mathbf{b}}(\delta_t; x_t^{(2)}, F)$ is identified for $\mathbf{b} \in \mathbf{C}_1$ and $(\delta_t, x_t^{(2)}) \in \mathbb{R}^{C_1} \times \mathbf{D}_x^{(2)}$. Steps 1 and 2 prove the first statement. The proof of the second statement is a combination of that of the first statement and a direct application of Theorem 3 of Wang (2020). We prove steps 1 and 2 for any given $x_{t\mathbf{J}}^{(2)} \in \mathbf{D}_x^{(2)}$. To simplify the exposition, I drop $x_{t\mathbf{J}}^{(2)}$ from the notation. Throughout the proof, denote the Fourier transformation of function ϕ by $\mathscr{F}(\phi)$.

Suppose that there exist (Γ', F') such that

$$s_{j\cdot}(\delta_{\mathbf{J}}; \Gamma', F') = s_{j\cdot}(\delta_{\mathbf{J}}; \Gamma, F)$$

for any $j \in \mathbf{J}$ and $\delta \in \mathbb{R}^J$.

Step 1: $\Gamma' = \Gamma$**.** Without loss of generality, we show $\Gamma_{(1,2)} = \Gamma'_{(1,2)}$. First, note that by letting all δ'_{tl}'s, $l \neq 1, 2$ tend to $-\infty$, market shares of single products l and of all bundles that contain any product $l \neq 1, 2$ converge to zero. Consequently, we obtain:

$$\begin{aligned} s_{1\cdot}(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F') &= s_{1\cdot}(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F), \\ s_{2\cdot}(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F') &= s_{2\cdot}(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F), \end{aligned} \tag{E.1}$$

for any $\delta'_{t\{1,2\}} = (\delta'_{t1}, \delta'_{t2}) \in \mathbb{R}^2$. Take the first equation in (E.1) and compute the partial derivatives with respect to δ'_{t2}:

$$\begin{aligned} \frac{\partial s_{1\cdot}(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F')}{\partial \delta'_{t2}} &= \int \frac{(e^{\delta'_{t1}+\delta'_{t2}+\mu_{t1}(\theta_{it})+\mu_{t2}(\theta_{it})})(e^{\Gamma'_{(1,2)}} - 1)}{(1 + e^{\delta'_{t1}+\mu_{t1}(\theta_{it})} + e^{\delta'_{t2}+\mu_{t2}(\theta_{it})} + e^{\delta'_{t1}+\mu_{t1}(\theta_{it})+\delta'_{t1}+\mu_{t1}(\theta_{it})+\Gamma'_{(1,2)}})^2} dF'(\theta_{it}), \\ \frac{\partial s_{1\cdot}(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F)}{\partial \delta'_{t2}} &= \int \frac{(e^{\delta'_{t1}+\delta'_{t2}+\mu_{t1}(\theta_{it})+\mu_{t2}(\theta_{it})})(e^{\Gamma_{(1,2)}} - 1)}{(1 + e^{\delta'_{t1}+\mu_{t1}(\theta_{it})} + e^{\delta'_{t2}+\mu_{t2}(\theta_{it})} + e^{\delta'_{t1}+\mu_{t1}(\theta_{it})+\delta'_{t1}+\mu_{t1}(\theta_{it})+\Gamma_{(1,2)}})^2} dF(\theta_{it}), \\ \frac{\partial s_{1\cdot}(\delta'_{t\{1,2\}}; \Gamma'_{(1,2)}, F')}{\partial \delta'_{t2}} &= \frac{\partial s_{1\cdot}(\delta'_{t\{1,2\}}; \Gamma_{(1,2)}, F)}{\partial \delta'_{t2}}. \end{aligned}$$

I can then cancel out $e^{\delta'_{t1}+\delta'_{t2}}$ in the nominators of $\frac{\partial s_{1\cdot}(\delta'_{t\{1,2\}};\Gamma'_{(1,2)},F')}{\partial \delta'_{t2}}$ and $\frac{\partial s_{1\cdot}(\delta'_{t\{1,2\}};\Gamma'_{(1,2)},F')}{\partial \delta'_{t2}}$. Letting $\delta_{t2} \to -\infty$, I obtain:

$$[e^{\Gamma'_{(1,2)}} - 1] \int \frac{e^{\mu_{t1}(\theta_{it})+\mu_{t2}(\theta_{it})}}{(1 + e^{\delta'_{t1}+\mu_{t1}(\theta_{it})})^2} dF'(\theta_{it}) = [e^{\Gamma_{(1,2)}} - 1] \int \frac{e^{\mu_{t1}(\theta_{it})+\mu_{t2}(\theta_{it})}}{(1 + e^{\delta'_{t1}+\mu_{t1}(\theta_{it})})^2} dF(\theta_{it}). \tag{E.2}$$

From (E.2), if $\Gamma_{(1,2)} = 0$, then $\Gamma'_{(1,2)} = \Gamma_{(1,2)} = 0$.Suppose $\Gamma_{(1,2)} \neq 0$. Denote the density functions of $\mu_{it} = (\mu_{it1}, \mu_{it2}) = (\mu_{t1}(\theta_{it}), \mu_{t2}(\theta_{it}))$ for $\theta_{it} \sim F$ and $\theta_{it} \sim F'$ by f_μ and f'_μ, respectively. Then, I can re-write (E.2) as:

$$[e^{\Gamma'_{(1,2)}} - 1] \int \frac{e^{\mu_{it1}+\mu_{it2}}}{(1 + e^{\mu_{it1}})^2} f'_\mu(\mu_{it1} - \delta'_{t1}, \mu_{it2}) d\mu_{it} = [e^{\Gamma_{(1,2)}} - 1] \int \frac{e^{\mu_{it1}+\mu_{it2}}}{(1 + e^{\mu_{it1}})^2} f_\mu(\mu_{it1} - \delta'_{t1}, \mu_{it2}) d\mu_{it}.$$

Define $g(\lambda) = \frac{e^\lambda}{(1+e^\lambda)^2}$. Then,

$$[e^{\Gamma'_{(1,2)}} - 1] \int g(\mu_{it1}) \tilde{f}'_\mu(\mu_{it1} - \delta'_{t1}) d\mu_{it1} = [e^{\Gamma_{(1,2)}} - 1] \int g(\mu_{it1}) \tilde{f}_\mu(\mu_{it1} - \delta'_{t1}) d\mu_{it1}, \tag{E.3}$$

where $\tilde{f}'_\mu(\mu_{it1}) = \int e^{\mu_{it2}} f'_\mu(\mu_{it1}, \mu_{it2}) d\mu_{it2}$ and $\tilde{f}_\mu(\mu_{it1}) = \int e^{\mu_{it2}} f_\mu(\mu_{it1}, \mu_{it2}) d\mu_{it2}$. Either side of (E.3) defines a convolution. Note that $g(\cdot), \tilde{f}'_\mu, \tilde{f}_\mu \in L^1(\mathbb{R})$. Consequently, I apply Fourier transformation on both sides of (E.3) and obtain:

$$[e^{\Gamma'_{(1,2)}} - 1] \mathscr{F}(g)(t) \mathscr{F}(\tilde{f}'_\mu)(t) = [e^{\Gamma_{(1,2)}} - 1] \mathscr{F}(g)(t) \mathscr{F}(\tilde{f}_\mu)(t),$$

for any $t \in \mathbb{R}$. Particularly, at $t = 0$, $\mathscr{F}(g)(0) > 0$. Then,

$$[e^{\Gamma'_{(1,2)}} - 1]\mathscr{F}(\tilde{f}'_\mu)(0) = [e^{\Gamma_{(1,2)}} - 1]\mathscr{F}(\tilde{f}_\mu)(0), \tag{E.4}$$

Note that $\mathscr{F}(\tilde{f}'_\mu)(t) = E_{\tilde{f}'}[e^{\mu_{it2}}]$ and $\mathscr{F}(\tilde{f}_\mu)(t) = E_{\tilde{f}}[e^{\mu_{it2}}]$. If they are equal, then $\Gamma'_{(1,2)} = \Gamma_{(1,2)}$. In particular, if $\Gamma_{(1,2)} = -\infty$, i.e. bundle $(1, 2)$ is not in the choice set, I obtain that $\Gamma'_{(1,2)} = -\infty$ and therefore identify that bundle $(1, 2)$ is not in the choice set. In what follows, we prove $E_{\tilde{f}'}[e^{\mu_{it2}}] = E_{\tilde{f}}[e^{\mu_{it2}}]$.

Take the second equation of (E.1) and let $\delta'_{it1} \to -\infty$. I then obtain:

$$\int \frac{e^{\delta'_{t2}+\mu_{it2}}}{1 + e^{\delta'_{t2}+\mu_{it2}}} f'_\mu(\mu_{it}) = \int \frac{e^{\delta'_{t2}+\mu_{it2}}}{1 + e^{\delta'_{t2}+\mu_{it2}}} f_\mu(\mu_{it}). \tag{E.5}$$

I cancel out $e^{\delta'_{t2}}$ from the nominators on both sides of (E.5) and let $\delta'_{2t} \to -\infty$. I then obtain $\int e^{\mu_{it2}} f'_\mu(\mu_{it}) = \int e^{\mu_{it2}} f_\mu(\mu_{it})$, i.e. $E_{\tilde{f}'}[e^{\mu_{it2}}] = E_{\tilde{f}}[e^{\mu_{it2}}]$.

Step 2: $s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; F') = s_{\mathbf{b}}(\delta_{\mathbf{C}_1}; F)$ for any $\mathbf{b} \in \mathbf{C}_1$ and $\delta_{\mathbf{C}_1} \in \mathbb{R}^{C_1}$. I prove this result for for the model of demand for multiple products within category. The proof is similar for the model of demand for multiple products across two categories.

Recall that the density function of $\mu_{it\mathbf{J}} = \mu_{t\mathbf{J}}(\theta_{it})$ for $\theta_{it} \sim F'$ and $\theta_{it} \sim F$ are f'_μ and f_μ, respectively. It suffices to prove that $f'_\mu = f_\mu$ almost everywhere. In the model of demand for multiple products within category in Theorem 3, plug $\Gamma' = \Gamma$ into the product-level market share function of j. I then have for any $\delta'_{t\mathbf{J}} \in \mathbb{R}^J$, $s_{j\cdot}(\delta'_{t\mathbf{J}}; \Gamma, F) = s_{j\cdot}(\delta'_{t\mathbf{J}}; \Gamma, F')$. According to the arguments in Appendix 8.13 of Iaria and Wang (2019), given the product-level market share functions and Γ, one can uniquely determine the bundle-level market shares, as function of $\delta_{t\mathbf{J}}$. Because both the product-level market share functions and Γ are identified, then $s_{\mathbf{b}}(\delta_t(\Gamma); F)$, where $\delta_t(\Gamma) = (\delta_{t1}, ..., \delta_{tJ}, (\delta_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2} + \Gamma)$, is identified as a function of $\delta_{t\mathbf{J}}$, for any $\mathbf{b} \in \mathbf{C}_1$. Consequently, the market share function of the outside option, $s_0(\delta_t(\Gamma); F)$, is identified as a function of $\delta_{t\mathbf{J}}$: for any $\delta'_{t\mathbf{J}} \in \mathbb{R}^J$

$$s_0(\delta'_{t\mathbf{J}}(\Gamma); F) = s_0(\delta'_{t\mathbf{J}}(\Gamma); F'). \tag{E.6}$$

I compute the higher-order cross derivative of both sides of (E.6):

$$\begin{aligned}
\frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F)}{\partial\delta'_{t1}, ..., \partial\delta'_{tJ}} &= \int \frac{P_\Gamma(\delta_{t1} + \mu_{it1}, ..., \delta_{tJ} + \mu_{itJ})}{Q_\Gamma(\delta_{t1} + \mu_{it1}, ..., \delta_{tJ} + \mu_{itJ})} f_\mu(\mu_{it}) d\mu_{it}, \\
\frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F')}{\partial\delta'_{t1}, ..., \partial\delta'_{tJ}} &= \int \frac{P_\Gamma(\delta_{t1} + \mu_{it1}, ..., \delta_{tJ} + \mu_{itJ})}{Q_\Gamma(\delta_{t1} + \mu_{it1}, ..., \delta_{tJ} + \mu_{itJ})} f'_\mu(\mu_{it}) d\mu_{it}, \\
\frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F)}{\partial\delta'_{t1}, ..., \partial\delta'_{tJ}} &= \frac{\partial^J s_0(\delta'_{t\mathbf{J}}(\Gamma); F')}{\partial\delta'_{t1}, ..., \partial\delta'_{tJ}},
\end{aligned} \tag{E.7}$$

where

$$Q_\Gamma(u_{it1}, ..., u_{itJ}) = 1 + \sum_{j\in\mathbf{J}} e^{u_{itj}} + \sum_{j<j'} e^{\Gamma_{(j,j')}} e^{u_{itj}+u_{itj'}}.$$

and

$$\begin{aligned}
P_\Gamma(u_{it1}, ..., u_{itJ}) &= \sum_{S\in\mathbf{S}} A(S) \prod_{(j,j')\in S} \frac{e^{u_{it(j,j')}}}{Q_\Gamma(u_{it1}, ..., u_{itJ})} \prod_{j\in S} \frac{\sum_{\mathbf{b}:\mathbf{b}\ni j} e^{u_{it\mathbf{b}}}}{Q_\Gamma(u_{it1}, ..., u_{itJ})} \\
&= \prod_{j\in\mathbf{J}} e^{u_{itj}} \sum_{S\in\mathbf{S}} A(S) \prod_{(j,j')\in S} \frac{e^{\Gamma_{(j,j')}}}{Q_\Gamma(u_{it1}, ..., u_{itJ})} \prod_{j\in S} \frac{1 + \sum_{j'\neq j} e^{u_{itj'}+\Gamma_{(j,j')}}}{Q_\Gamma(u_{it1}, ..., u_{itJ})}
\end{aligned} \tag{E.8}$$

where S is a partition of $\{1, ..., J\}$ with each part being at most size 2, $\mathbf{S}$ collects all such partitions which are the results of the higher-order cross derivative $\frac{\partial^J}{\partial\delta'_{t1},...,\partial\delta'_{tJ}}$, and $A(S)$ is a constant depending on the partition $S \in \mathbf{S}$. An example of S is $\{\{1\}, \{2, 5\}, \{4\}, \{3, 6\}\}$. Each term in the products of P_Γ corresponds to the choice probability of either bundle (j, j') or the product-level choice probability of product j, evaluated at $u_{it\mathbf{J}}$ and Γ, and bounded by 1. From (E.7), I obtain:

$$\int \frac{P_\Gamma((e^{\lambda_{itj}})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_{itj}})_{j\in\mathbf{J}})} [f_\mu(\lambda_{it\mathbf{J}} - \delta'_{t\mathbf{J}}) - f_\mu(\lambda_{it\mathbf{J}} - \delta'_{t\mathbf{J}})] d\lambda_{it\mathbf{J}} = 0, \tag{E.9}$$

for any $\delta'_{t\mathbf{J}} \in \mathbb{R}^J$. I prove the following lemma:

Lemma 3.

- $\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})} \in L^1(\mathbb{R}^J)$.
- *The zero set of* $\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)$ *in* $\mathbb{R}^J$ *is of zero Lebesgue measure.*

Note that the right-hand side of (E.9) is a convolution. Because of the first statement of Lemma 3, I can apply Fourier transformation on both sides and obtain:

$$\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right) \mathscr{F}(f_\mu - f'_\mu) = 0.$$

Applying the second statement of Lemma 3, I obtain $\mathscr{F}(f_\mu) = \mathscr{F}(f'_\mu)$ almost everywhere. Due to the continuity of characteristics functions, $\mathscr{F}(f_\mu) = \mathscr{F}(f'_\mu)$ everywhere and hence the distribution of μ_{it} is identified. In the remaining part, I prove Lemma 3.

Proof. First, we make the transformation of variables $\lambda_\mathbf{J}$ to $e^{\lambda_\mathbf{J}}$:

$$\int \left| \frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})} \right| d\lambda_\mathbf{J} = \int_{\mathbb{R}^J_+} \left| \frac{P_\Gamma(y_1, ..., y_J)}{Q_\Gamma(y_1, ..., y_J)} \frac{1}{\prod_{j=1}^J y_j} \right| dy_\mathbf{J}.$$

For the first statement, because of (E.8), it suffices to prove that for each $S \in \mathbf{S}$,

$$\left| \frac{1}{Q_\Gamma(y_1, ..., y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \right| \tag{E.10}$$

is integrable in $\mathbb{R}_+^J$. To show this, I divide $\mathbb{R}_+^J$ into 2^J regions: $\mathbb{R}_+^J = \times_{j=1}^J I_j$, where $I_j = (0,1], (1, +\infty)$. Then, it is enough to prove that (E.10) is integrable in each of these regions. Without loss of generality, suppose that the region is $R_k = \{(y_1, ..., y_J) : y_j \in (0,1), j = 1, ..., k;\ y_{j'} \geq 1, j' = k+1, ..., J\}$. Then, for a given j, we have four cases to control:

1. $j \leq k$ and j appears in S as (j, j').
2. $j \leq k$ and j appears in S as j.
3. $j > k$ and j appears in S as (j, j').
4. $j > k$ and j appears in S as j.

Note that for cases 1 and 2, the corresponding terms in (E.10) can be controlled by e^{Γ_m} with $\Gamma_m = \max\{0, (\Gamma_{(j,j')})_{j \leq j'}\}$. For case 3, $\frac{e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1,...,y_J)} \leq \frac{e^{\Gamma_m}}{y_j}$. For case 4,

$$\frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \leq \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma_{(j,j')}}}{y_j + \sum_{j' \neq j} y_j y_{j'} e^{\Gamma_{(j,j')}}} = \frac{1}{y_j} \leq \frac{e^{\Gamma_m}}{y_j}.$$

Moreover,

$$\begin{aligned} Q_\Gamma(y_1, ..., y_J) &\leq \frac{1}{\sum_{j>k} y_j + \sum_{k<j<j'} y_j y_{j'} e^{\Gamma_{(j,j')}}} \\ &\leq \frac{2}{(J-k)(J-k+1)(\prod_{k<j<j'} e^{\Gamma_{(j,j')}})^{\frac{2}{(J-k)(J-k+1)}} \prod_{j=k+1}^J y_j^{\frac{2}{J-k+1}}}. \end{aligned}$$

The last step is due to the inequality of arithmetic and the geometric means. Then, for all the four cases, we have:

$$\left| \frac{1}{Q_\Gamma(y_1, ..., y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \right| \leq A(J, k, \Gamma) \prod_{j=k+1}^{J} y_j^{-1-\frac{2}{J-k+1}}, \tag{E.11}$$

where $A(J, k, \Gamma) = \frac{2e^{J\Gamma_m}}{(J-k)(J-k+1)(\prod_{k<j<j'} e^{\Gamma_{(j,j')}})^{\frac{2}{(J-k)(J-k+1)}}}$. Note that $\prod_{j=k+1}^J y_j^{-1-\frac{2}{J-k+1}}$ is integrable in R_k and $A(J, k, \Gamma)$ is a constant. Then,

$$\left| \frac{1}{Q_\Gamma(y_1, ..., y_J)} \prod_{(j,j') \in S} \frac{e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \prod_{j \in S} \frac{1 + \sum_{j' \neq j} y_{j'} e^{\Gamma_{(j,j')}}}{Q_\Gamma(y_1, ..., y_J)} \right|$$

is integrable in $\{(y_1, ..., y_J) : y_j \in (0,1), j = 1, ..., k;\ y_{j'} \geq 1, j' = k+1, ..., J\}$. The proof of the first statement is completed.

To prove the second statement, according to Mityagin (2015), it suffices to show that the real (or imaginary) part of $\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)$ is non-constant real analytic function. In order to prove the real analytic property, the key is to control the higher order derivatives of $\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)(y)$:

$$\frac{\partial^L \mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)(y)}{\prod_{j=1}^J \partial y_j^{l_j}} = \mathscr{F}\left(\prod_{j=1}^J (-\mathrm{i}\lambda_j)^{l_j} \frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)(y), \tag{E.12}$$

where $\sum_{j=1}^J l_j = L$ and i is the imaginary unit. I now prove that this higher order derivative can be controlled by $\left(\frac{J+1}{2}\right)^L \prod_{j=1}^J l_j!$. This result will then imply that for any $y \in \mathbb{R}^J$, there exist $0 < \epsilon < \frac{2}{J+1}$ such that for $y' \in \mathbb{R}^J$ and $|y' - y| < \epsilon$, the Taylor expansion of $\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)(y')$ around y uniformly converges to $\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)(y')$. Consequently, $\mathscr{F}\left(\frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right)(y)$ is everywhere real analytic in $\mathbb{R}^J$. It is not constantly zero because $\frac{P_\Gamma}{Q_\Gamma}$ is not constantly zero. In the remaing part of the proof, I prove (E.12) can be controlled by $\left(\frac{J+1}{2}\right)^L \prod_{j=1}^J l_j!$.
It suffices to study $\int \left|\prod_{j=1}^J (|\lambda_j|)^{l_j} \frac{P_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}{Q_\Gamma((e^{\lambda_j})_{j\in\mathbf{J}})}\right| d\lambda$, or equivalently,

$$\int_{\mathbb{R}_+^J} \left|\prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{P_\Gamma((y_j)_{j\in\mathbf{J}})}{Q_\Gamma((y_j)_{j\in\mathbf{J}})} \frac{1}{\prod_{j=1}^J y_j}\right| dy_{\mathbf{J}}.$$

I follow the same technique as in the proof of the first statement and evaluate, for each $S \in \mathbf{S}$,

$$\left|\prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{1}{Q_\Gamma(y_1, ..., y_J)} \prod_{(j,j')\in S} \frac{e^{\Gamma(j,j')}}{Q_\Gamma(y_1, ..., y_J)} \prod_{j\in S} \frac{1 + \sum_{j'\neq j} y_{j'} e^{\Gamma(j,j')}}{Q_\Gamma(y_1, ..., y_J)}\right| \tag{E.13}$$

in each of the 2^J regions. Without loss of generality, for region R_k, using (E.11), we have:

$$\begin{aligned}
&\left|\prod_{j=1}^J (|\ln y_j|)^{l_j} \frac{1}{Q_\Gamma(y_1, ..., y_J)} \prod_{(j,j')\in S} \frac{e^{\Gamma(j,j')}}{Q_\Gamma(y_1, ..., y_J)} \prod_{j\in S} \frac{1 + \sum_{j'\neq j} y_{j'} e^{\Gamma(j,j')}}{Q_\Gamma(y_1, ..., y_J)}\right| \\
&\leq A(J,k,\Gamma) \prod_{j=1}^k |\ln y_j|^{l_j} \prod_{j=k+1}^J |\ln y_j|^{l_j} y_j^{-1-\frac{2}{J-k+1}}.
\end{aligned}$$

Finally,

$$\int_0^1, ..., \int_0^1 \int_1^\infty ... \int_1^\infty A(J,k,\Gamma) \prod_{j=1}^{k} |\ln y_j|^{l_j} \prod_{j=k+1}^{J} |\ln y_j|^{l_j} y_j^{-1-\frac{2}{J-k+1}} dy_{\mathbf{J}}$$
$$= A(J,k,\Gamma)(\frac{J-k+1}{2})^{L+J-k} \prod_{j=1}^{J} l_j!$$
$$\leq A(J,k,\Gamma)(\frac{J+1}{2})^{J}(\frac{J+1}{2})^{L} \prod_{j=1}^{J} l_j!$$

Consequently, when I sum over all the integrals in the 2^J regions and all $S \in \mathbf{S}$,

$$\int_{\mathbb{R}_+^J} \left| \prod_{j=1}^{J} (|\ln y_j|)^{l_j} \frac{P_\Gamma((y_j)_{j\in\mathbf{J}})}{Q_\Gamma((y_j)_{j\in\mathbf{J}})} \frac{1}{\prod_{j=1}^{J} y_j} \right| dy_{\mathbf{J}}$$

will be bounded by $(\frac{J+1}{2})^L \prod_{j=1}^{J} l_j!$ multiplied by some constant only depending on J and Γ. The proof is completed. $\square$

F Proof of Corollary 1

I will construct (Γ_0, F_0') and (Γ_0', F_0) such that $\Gamma_0 \neq \Gamma_0'$ and $F_0 \neq F_0'$, while $s_{1\cdot}(\cdot;\Gamma_0,F_0') = s_{1\cdot}(\cdot;\Gamma_0',F_0)$. Because $F_0 \neq F_0'$, then $s_{(1,1)}(\cdot;F_0) \neq s_{(1,1)}(\cdot;F_0')$.

First, I compute the derivative of $s_{1\cdot}(\delta;\Gamma,F)$ with respect to δ:

$$\begin{aligned} \frac{\partial s_{1\cdot}(\delta;\Gamma,F)}{\partial\delta} &= \int \frac{e^{\delta+\mu} + 4e^{2\delta+2\mu+\Gamma}}{(1+e^{\delta+\mu}+e^{2\delta+2\mu+\Gamma})^2} dF(\mu) \\ &= \int R(\delta+\mu;\Gamma) dF(\mu), \end{aligned} \tag{F.1}$$

where $R(x;\Gamma) = \frac{e^x+4e^{2x+\Gamma}}{(1+e^x+e^{2x+\Gamma})^2}$. Note that $R(\cdot;\Gamma) \in L^1(\mathbb{R})$. Define $\gamma = e^\Gamma$ and

$$V(\gamma) = \int_{\mathbb{R}} R(x;\Gamma)dx = \int_{\mathbb{R}^+} \frac{1+4\gamma t}{(1+t+\gamma t^2)^2} dt$$

$V(\gamma)$ is a continuous function of $\gamma \in [0,\infty)$, with $V(0) = 1 > V(\infty) = 0$. Moreover, $\lim_{\gamma\to 0^+} \frac{dV}{d\gamma} = +\infty > 0$. As a consequence, there exist $\gamma_0 \neq \gamma_0'$ and $\gamma_0, \gamma_0' > 0$, such that $V(\gamma_0) = V(\gamma_0')$. Therefore, there exists $\Gamma_0 = \ln\gamma_0 > -\infty$ and $\Gamma_0' = \ln\gamma_0' > -\infty$, such that $\Gamma_0 \neq \Gamma_0'$ and $V_0 = \int_{\mathbb{R}} R(x;\Gamma_0) = \int_{\mathbb{R}} R(x;\Gamma_0')dx$. Note that $\frac{R(\cdot;\Gamma_0)}{V_0}$ and $\frac{R(\cdot;\Gamma_0')}{V_0}$ are both well-defined but different density functions. Denote the corresponding distribution functions by F_0 and F_0', respectively: $\frac{dF_0}{d\mu} = \frac{R(\mu;\Gamma_0)}{V_0}$ and $\frac{dF_0'}{d\mu} = \frac{R(\mu;\Gamma_0')}{V_0}$.

Based on (F.1), consider the Fourier transformation of $\frac{\partial s_{1.}(\delta;\Gamma_0,F_0')}{\partial\delta}$ and $\frac{\partial s_{1.}(\delta;\Gamma_0',F_0)}{\partial\delta}$:

$$\begin{aligned}\mathscr{F}\left(\frac{\partial s_{1.}(\delta;\Gamma_0,F_0')}{\partial\delta}\right)(t) &= \mathscr{F}\left(R(\cdot;\Gamma_0)\right)(t)\mathscr{F}\left(\frac{dF_0'}{d\mu}\right)(t),\\ \mathscr{F}\left(\frac{\partial s_{1.}(\delta;\Gamma_0',F_0)}{\partial\delta}\right)(t) &= \mathscr{F}(R(\cdot;\Gamma_0'))(t)\mathscr{F}\left(\frac{dF_0}{d\mu}\right)(t).\end{aligned}$$

Then,

$$\mathscr{F}\left(\frac{\partial s_{1.}(\delta;\Gamma_0,F_0')}{\partial\delta}\right)(t) = \mathscr{F}\left(\frac{\partial s_{1.}(\delta;\Gamma_0',F_0)}{\partial\delta}\right)(t),$$

Consequently, $\frac{\partial s_{1.}(\delta;\Gamma_0,F_0')}{\partial\delta} = \frac{\partial s_{1.}(\delta;\Gamma_0',F_0)}{\partial\delta}$ for $\delta\in\mathbb{R}$, and $s_{1.}(\delta;\Gamma_0',F_0) - s_{1.}(\delta;\Gamma_0,F_0')$ is a constant function in $\mathbb{R}$. Taking $\delta=+\infty$, we obtain that this constant is zero and hence $s_{1.}(\delta;\Gamma_0',F_0) = s_{1.}(\delta;\Gamma_0,F_0')$ for $\delta\in\mathbb{R}$. The construction is completed.

G Proof of Proposition 2

The proof directly follows from that of Lemma 2.

H Construction of Product-Level Market Shares

In this appendix, I illustrate the construction of product-level market shares from market-level sales data. I suppress t to simplify the exposition. Suppose that there are I households and the size of household $i=1,...,I$ is $n_i\in\{1,...,N\}$. Denote by q_k the weekly per capita consumption of the relevant products of category k (breakfast cereals or milk). Then, for product j in category k, the total consumption D_{jk}, is:

$$\begin{aligned}D_{jk} &= \sum_{i=1}^{I}\sum_{\mathbf{b}:\mathbf{b}\ni j}\mathbf{1}\{i\text{ buys }\mathbf{b}\}n_i q_k\\ &= q_k\sum_{n=1}^{N} n\sum_{\mathbf{b}:\mathbf{b}\ni j}\sum_{i=1}^{I}\mathbf{1}\{i\text{ buys }\mathbf{b}, n_i=n\}\\ &= Iq_k\sum_{n=1}^{N} n\sum_{\mathbf{b}:\mathbf{b}\ni j}\frac{\sum_{i=1}^{I}\mathbf{1}\{i\text{ buys }\mathbf{b}, n_i=n\}}{I}.\end{aligned}$$

Denote by $\mathscr{s}_{\mathbf{b}}^{n}$ the average purchase probability of bundle $\mathbf{b}$ among households of size n. Then, when I is very large,

$$\begin{aligned}
\frac{D_{jk}}{Iq_k} &= \sum_{n=1}^{N} n \sum_{\mathbf{b}:\mathbf{b}\ni j} \frac{\sum_{i=1}^{I} \mathbf{1}\{i \text{ buys } \mathbf{b}, n_i = n\}}{I} \\
&\approx \sum_{n=1}^{N} n\pi_n \sum_{\mathbf{b}:\mathbf{b}\ni j} \mathscr{s}_{\mathbf{b}}^{n} \\
&= \sum_{n=1}^{N} n\pi_n \mathscr{s}_{j\cdot}^{n}, \\
\frac{D_{jk}}{I\bar{N}q_k} &\approx \sum_{n=1}^{N} \bar{\pi}_n \mathscr{s}_{j\cdot}^{n} = \mathscr{s}_{j\cdot},
\end{aligned}$$

where $\bar{N} = \sum_{n=1}^{N} n\pi_n$ is the average household size and $\{\bar{\pi}_n\}_{n=1}^{N}$ is the distribution of household sizes weighted by size. Note that when computing product-level market shares, one should use the weighted distribution $\{\bar{\pi}_n\}_{n=1}^{N}$ rather than $\{\pi_n\}_{n=1}^{N}$ to properly take into account heterogeneous consumption across households of different sizes.

Under the assumptions in section 6, D_{jk} is equal to the sales in lbs of product j of category k. Moreover, the IRI dataset contains information the sampled households with which we can infer the number of households I and the distribution of their demographics. Finally, for d_k, $k \in \mathbf{K}$, I use external sources: the weekly per capita consumption of breakfast cereals is 0.19 lbs and that of fluid milk is 3.4 lbs.[52] Based on these pieces of information, we construct $\mathscr{s}_{j\cdot}$'s.[53]

I Main Tables

[52] See https://hypertextbook.com/facts/2006/LauraFalci.shtml for a collection of these reports.

[53] Another implicit assumption is that the ratio of consumption between breakfast cereals and fluid milk is the same across households of different sizes. However, it is possible that this is not true. For example, households with children may consume relatively more fluid milk.

Table 9: RTE Cereal Products

Brand	Flavour	Fortification	Grain
General Mills Cherrios	Toasted	Missing	WHOLE GRAIN OAT
General Mills Cinnamon TST CR	Cinnamon Toast	12 ESSNTL VTMN&MNRL	WHOLE WHEAT AND RICE
General Mills Cinnamon TST CR	Cinnamon Toast	Missing	WHOLE WHEAT AND RICE
General Mills Honey Nut Cheer	Honey Nut	Missing	WHL GRAIN OAT & BRLY
General Mills Honey Nut Cheer	Honey Nut	Missing	WHOLE GRAIN OAT
General Mills Lucky Charms	Toasted	CALCIUM & VITAMIN D	WHOLE GRAIN OAT
General Mills Lucky Charms	Toasted	Missing	WHOLE GRAIN OAT
General Mills Multi Grain Che.	Regular	10 VITAMINS&MINERALS	MULTI GRAIN
Kashi Go Lean Crunch	Regular	Missing	MULTI GRAIN
Kellogg's Apple Jacks	Apple Cinnamon	Missing	3 GRAIN
Kellogg's Corn Flakes	Regular	Missing	CORN
Kellogg's Frosted Flakes	Regular	VITAMIN D	CORN
Kellogg's Frosted Mini Wheats	Regular	Missing	WHOLE GRAIN WHEAT
Kellogg's Raisin Bran	Regular	Missing	WHL GRN WHT WHT BRN
Kellogg's Rice Krispies	Toasted	Missing	RICE
Kellogg's Special K	Toasted	Missing	RICE
Kellogg's Special K Fruit & Yo	Regular	Missing	OAT RICE WHEAT
Kellogg's Special K Red Berrie	Regular	Missing	RICE AND WHEAT
Kellogg's Special K Vanilla AL	Regular	Missing	RICE AND WHEAT
Post Grape Nuts	Regular	Missing	WHOLE GRN WHT & BRLY
Post Honey Bunches of Oats	Honey	Missing	WHOLE GRAIN OAT
Post Honey Bunches of Oats	Honey Roasted	Missing	WHOLE GRAIN OAT
Post Raisin Bran	Regular	Missing	WHOLE GRAN WHT & BRN
Post Selects Great Grains	Regular	Missing	MULTI GRAIN
Private Label	Regular	Missing	GRANOLA

Table 10: Milk Products

Brand	Flavour	Fortification	Fat Content	Type of Milk
GARELICK FARMS	WHITE	VITAMIN A & D	skimmed	dairy
GARELICK FARMS	WHITE	VITAMIN A & D	low fat	dairy
GARELICK FARMS	WHITE	VITAMIN D	whole fat	dairy
GARELICK FARMS TRUMOO	CHOCOLATE	MISSING	low fat	dairy
GUIDAS	WHITE	VITAMIN A & D	skimmed	dairy
GUIDAS	WHITE	VITAMIN A & D	low fat	dairy
GUIDAS	WHITE	VITAMIN D	whole fat	dairy
HIGH LAWN FARM	WHITE	VITAMIN A & D	whole fat	dairy
HIGH LAWN FARM	WHITE	VITAMIN A & D	skimmed	dairy
HIGH LAWN FARM	WHITE	VITAMIN A & D	low fat	dairy
HOOD	WHITE	VITAMIN A C D W CLCM	skimmed	dairy
HOOD	WHITE	VITAMIN A C D W CLCM	low fat	dairy
HOOD	WHITE	VIT C D CALCIUM	whole fat	dairy
HOOD LACTAID	WHITE	VITAMIN A & D	low fat	dairy
HOOD SIMPLY SMART	WHITE	VIT A & D W/CALC&PROTN	skimmed	dairy
PRIVATE LABEL	CHOCOLATE	MISSING	low fat	dairy
PRIVATE LABEL	WHITE	VITAMIN A & D	skimmed	dairy
PRIVATE LABEL	WHITE	VITAMIN A & D W/CALC	skimmed	dairy
PRIVATE LABEL	WHITE	VITAMIN A & D	low fat	dairy
PRIVATE LABEL	WHITE	VITAMIN D	whole fat	dairy

Table 11: Average Estimated Own- and Cross-Price Elasticities (Model II), Grain Type and Fat Content

	RTE cereals			Milk		
	uni-grain	multi-grain	granola	skimmed	low fat	whole fat
RTE cereals, uni-grain	-1.512	0.231	0.009	-0.007	-0.017	-0.008
multi-grain	0.318	-1.642	0.010	-0.007	-0.018	-0.009
granola	0.214	0.185	-1.518	-0.009	-0.019	-0.009
Milk, skimmed	-0.078	-0.061	-0.005	-0.327	0.056	0.029
low fat	-0.079	-0.061	-0.005	0.022	-0.280	0.029
whole fat	-0.078	-0.061	-0.005	0.022	0.056	-0.319

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row characteristics with respect to a 1% increase in the prices of the products collected by the column characteristics.

Table 12: Average Estimated Own- and Cross-Price Elasticities (Model II), Flavours

	RTE cereals		Milk	
	unflavoured	flavoured	unflavoured	chocolate
RTE cereals, unflavoured	-1.504	0.179	-0.036	-0.0003
flavoured	0.137	-1.425	-0.028	-0.001
Milk, unflavoured	-0.071	-0.072	-0.235	0.002
chocolate	-0.036	-0.185	0.096	-0.366

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row characteristics with respect to a 1% increase in the prices of the products collected by the column characteristics.

Table 13: Average Estimated Own- and Cross-Price Elasticities (Model II), Fortification and Flavours

	RTE cereals		Milk	
	unfortified	fortified	unflavoured	chocolate
RTE cereals, unfortified	-1.355	0.070	-0.032	-0.001
fortified	0.454	-1.804	-0.033	-0.001
Milk, unflavoured	-0.123	-0.019	-0.235	0.002
chocolate	-0.197	-0.023	0.096	-0.366

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row characteristics with respect to a 1% increase in the prices of the products collected by the column characteristics.

Table 14: Average Estimated Own- and Cross-Price Elasticities (Model II), Producer

	RTE cereals					Milk				
	General Mills	Kashi	Kellogg's	Post	Private Label	Garelick Farms	Guidas	High Lawn Farm	Hood	Private Label
RTE cereals, General Mills	-1.556	0.009	0.188	0.052	0.008	-0.006	-0.026	-0.002	-0.002	-0.016
Kashi	0.151	-1.829	0.178	0.038	0.013	-0.007	-0.028	-0.003	-0.002	-0.020
Kellogg's	0.187	0.011	-1.572	0.038	0.011	-0.007	-0.031	-0.003	-0.002	-0.019
Post	0.222	0.011	0.176	-1.710	0.011	-0.007	-0.029	-0.003	-0.002	-0.020
Private Label	0.159	0.015	0.185	0.039	-1.518	-0.008	.	-0.003	-0.002	-0.023
Milk, Garelick Farms	-0.072	-0.006	-0.074	-0.018	-0.005	-0.356	.	0.006	0.006	0.059
Guidas	-0.029	-0.002	-0.033	-0.006	.	.	-0.189	0.013	0.002	0.015
High Lawn Farm	-0.059	-0.005	-0.064	-0.015	-0.005	0.017	0.132	-0.528	0.005	0.050
Hood	-0.058	-0.005	-0.063	-0.014	-0.005	0.017	0.126	0.007	-0.562	0.050
Private Label	-0.063	-0.005	-0.068	-0.016	-0.005	0.017	0.129	0.007	0.005	-0.273

Notes: Each entry reports the percent change in the sum of the product-level market shares of the products collected by the row producer with respect to a 1% increase in the prices of the products collected by the column producer.
The . refers to the situation where the products of the row and column brands are not available at the same time in any market in the data.

Online Appendix

J Data Generating Process in Section 3.3

In this appendix, I provide details on the DGP used in the counterfactual simulations in section 3.3. There are $t = 1, ..., T$ markets and $T = 1000$. In each market, we have two product categories $\mathbf{J}_1 = \{1, 2\}$ and $\mathbf{J}_2 = \{3, 4\}$. The indirect utility of individual i from purchasing product $j \in \mathbf{J}_1 \cup \mathbf{J}_2$ in market t:

$$U_{itj} = -2p_{tj} + x_{tj} + \xi_{tj} + \varepsilon_{itj},$$

the indirect utility from purchasing a bundle $\mathbf{b} = (j_1, j_2) \in \mathbf{J}_1 \times \mathbf{J}_2$ is

$$U_{it\mathbf{b}} = -2(p_{tj_1} + p_{tj_2}) + (x_{tj_1} + x_{tj_2}) + (\xi_{tj_1} + \xi_{tj_2}) + \Gamma_{j_1 j_2} + \varepsilon_{it\mathbf{b}},$$

and the indirect utility from purchasing the outside option 0 is $U_{it0} = \epsilon_{it0}$, where ε_{itj}, $\varepsilon_{it\mathbf{b}}$ and ε_{it0} are i.i.d. Gumbel, for $j \in \mathbf{J}_1 \cup \mathbf{J}_2$, $\mathbf{b} \in \mathbf{J}_1 \times \mathbf{J}_2$. Prices p_{tj} are generated from a Bertrand-price setting game in which the marginal production cost of product j in market t is positive and constant:

$$c_{tj} = 0.5 \exp\{0.5\} \exp\{z_{tj} + 0.2 w_{tj}\},$$

where z_{tj} is a product-market specific cost shifter and w_{tj} is a product-market specific supply shock. I solve the first-order conditions (FOCs) of the pricing game, from which I derive the price for each product in each market.[54] Finally, $(x_{tj}, z_{tj}, \xi_{tj}, w_{tj})$ are mutually independent and i.i.d. across j's and t's, with each being generated from a centred normal distribution with a variance 0.5. When estimating the model of demand with $\Gamma = 0$, I use (x_{tj}, z_{tj}) as instruments. When simulating the supermarket nonlinear pricing competition, I also assume that there is no additional cost of implement discounts, i.e. $c_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} c_{tj}$.

K Implementing (11) Using Jacobian-based Methods: Numerical Performance

In this appendix, I explore the numeric performance of Jacobian-based algorithms in implementing the demand inverse in (11). I compare convergence time across different algorithms and starting values—particularly, the proposed starting value in section 5.2—for various sizes of choice set.

[54]When solving the FOCs, I use the vector of marginal costs as starting value. The algorithm in Matlab returns an unique price vector.

Table 15 summarises the main results of the simulations. The DGP is a discrete choice model of bundles up to size two and the prices are generated from a Bertrand pricing game under complete information with constant marginal costs. I simulate 50 markets with the same structural parameters. The unobserved demand shocks $\xi_{t\mathbf{J}}$ are Gaussian and i.i.d. across markets and products. I implement the demand inverse using the true model and also using a demand model of single products.[55] I report the median convergence time (in seconds). Moreover, I compare the performances of these inverses using the starting value $\delta_*^{(0)}$ defined in (13) and the standard starting value $\delta^{(0)} = 0$. I replicate this setting for various sizes of choice set ($J = 10$, 50, and 100). For example, when $J = 100$, the true model has 5051 alternatives (100 single products, 4950 bundles of two different products, and an outside option). The demand inverse $s_{\mathbf{J}\cdot}^{-1}$ will then treat the observed market shares as those generated from the true model, while the demand inverse $s_{\mathbf{J}}^{-1}$ will treat the same observed market shares as if they are generated from a demand model of 100 single products.

Table 15: Demand Inverse of Product-Level Market Shares: Convergence Time

Algorithm	Trust-Region-Dogleg				Trust-Region-Reflective				Levenberg-Marquardt			
	$s_{\mathbf{J}\cdot}^{-1}$		$s_{\mathbf{J}}^{-1}$		$s_{\mathbf{J}\cdot}^{-1}$		$s_{\mathbf{J}}^{-1}$		$s_{\mathbf{J}\cdot}^{-1}$		$s_{\mathbf{J}}^{-1}$	
Init. Point	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0	$\delta_*^{(0)}$	0
# Products												
$J = 10$	0.04	0.09	0.03	0.08	0.05	0.09	0.04	0.08	0.08	0.09	0.07	0.09
50	0.49	1.45	0.10	2.64	0.41	1.31	0.13	0.34	1.31	1.92	0.19	0.12
100	4.50	12.22	0.12	3.21	3.32	12.15	0.27	0.60	12.25	20.21	0.33	0.18

Notes: Trust-region-dogleg, trust-region, and Levenberg-Marquardt algorithms are built-in algorithms of function *fsolve* in Matlab. All of them are large-scale and minimise the sum of squares of the components of (11). Median convergence time (in seconds) of 50 independently simulated markets is reported. Tolerance level in the stop criterion of all algorithms is set to 10^{-16}.

There are two main observations. First, for $s_{\mathbf{J}\cdot}^{-1}$, using the recommended starting value $\delta_*^{(0)}$ remarkably reduces convergence time in all cases. The gain is larger when the number of products is larger. When $J = 100$, trust-region-dogleg and trust-region-reflective algorithms reduce around 70% convergence time by using $\delta_*^{(0)}$ than using $\delta^{(0)} = 0$. The numerical gain for $s_{\mathbf{J}}^{-1}$ using $\delta_*^{(0)}$ is similar.[56] Second, as the problem size increases, using $\delta_*^{(0)}$ does not seem to increase the number of iterations required to converge. For example, the convergence time for $J = 100$ by

[55] The structural parameters are chosen so that the sum of the simulated product-level market shares is always smaller than one. This allows to implement the inverse of these simulated product-level market shares using models of demand for single products

[56] In this demand inverse, K in (13) is 1.

using the three Jacobian-based algorithms with $\delta_*^{(0)}$ is roughly 100 times of that for $J = 10$. Because the bundle size is at most two, then the size of choice set increases quadratically with respect to the number of products. Therefore, the number of required computations for one evaluation of market share functions also increases quadratically. While, the total convergence time when using $\delta_*^{(0)}$ seems to increase only quadratically for $s_{\mathbf{J}.}^{-1}$ with respect to the number of products, implying that the number of iterations does not increase as J increases.

L Identification of Product-Level Market Share Functions Using Other IVs.

In this Appendix, I develop similar identification arguments with other types of IVs. I will focus on BLP-type instruments and exogenous product characteristics.

BLP-type instruments Cost shifters are not always available to the econometrician. Moreover, the validity of Hausman-type instruments requires independence of demand shocks across markets of the same region or of the same time period. This can be violated whenever there is unobserved correlated demand shock across markets, such as national advertisement. In demand models of single products, Berry et al. (1995) proposed to use characteristics (and their functions) of other products in the same market as instruments. Their validity follows from the intuition that products with similar characteristics are closer substitutes. Then, "distance" in the space of product characteristics will be a good proxy of substitution among products.

Because such variables for product j, denoted as $x_{t,-j}$, are excluded from indirect utility of j, then, they can provide useful variation in price p_{tj} via the markup of product j that identifies (Γ, F). Formally, in (D.1), for the equation of product j, one can fix $x_{tj} = x_j$ and let $x_{t,-j} = (x_{tr})_{r \neq j}$ varies in $\mathbb{R}^{J-1}$.

It is worth noting that, different from cost shifters, BLP instruments may not always be able to provide useful variation even though they vary exogenously. For example, if prices p_{tj}'s are not responsive to $x_{t,-j}$, then there is no variation in p_{tj} due to the variation of $x_{t,-j}$. The unresponsiveness of prices with respect to BLP instruments can occur in a *large-market* setting when the number of products increases to infinity and therefore the competition between two products becomes very weak.[57] Asymptotically, product prices are no more functions of characteristics of other products, but only of their own characteristics. Then, BLP instruments (say, x_{tj} for product different from j) does not enter pricing functions of any other product ($p_{t,-j}$) and hence do not produce any exogenous variation in prices. In this paper, because I focus on *many-market* settings and the number of products is fixed (see Assumption 1(iii)), BLP instruments are still valid for the identification of the price coefficient, demand synergy parameters, and the distribution of the random coefficients.

L.1 Exogenous product characteristics

Suppose that prices are generated from a linear pricing simultaneous Bertrand game under complete information with constant marginal cost c_{tj} for $j \in \mathbf{J}$. I abstract from cost shifters in c_{tj} and denote the joint density function of $(\xi_{t\mathbf{J}}, c_{t\mathbf{J}})$ by $f_{\xi,c}$. I propose the following identification result:

Theorem L.1. *Suppose that Assumptions 1-3 and regularity condition 2 hold. If the following conditions hold:*

1. $x_{t\mathbf{J}}$ *is independent of* $(\xi_{t\mathbf{J}}, c_{t\mathbf{J}})$ *and the domain of* $x_{t\mathbf{J}}$ *is an open set in* $\mathbb{R}^{K_1}$.
2. $\alpha_i = \alpha \neq 0$, *and there exists a* k *such that* $\beta_{ik} = \beta_k$ *and* β_k/α *is identified.*
3. *Given* $x^{(2)}$, $p_{t\mathbf{J}} = p_{\mathbf{J}}(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}}; x^{(2)})$ *is a continuous function of* $(\beta x + \xi_{t\mathbf{J}}, c_{t\mathbf{J}})$.
4. F *has compact support.*

- *If* (ξ, c) *is Gaussian distributed, then* (α, β_k) *and* $s_{\mathbf{J}\cdot}(\cdot; x^{(2)}_{t\mathbf{J}}, \Gamma, F)$ *are identified.*
- *If the DGP is a model of demand for multiple products across categories, then under regularity condition 2 in Appendix D,* α *and* $s_{t\mathbf{J}\cdot}(\cdot; x^{(2)}, \Gamma, F)$ *are identified.*

The proof is similar to that of Theorem 2 and will proceed in three steps. I will skip similar parts and accent differences.

[57] Assuming that the distribution of random coefficients is priorly identified, Armstrong (2016b) provides conditions under which BLP instruments are weak for prices and therefore invalid for the identification of price coefficient. Intuitively, because prices and BLP instruments are correlated via markups, in his large-market setting, when the number of products increases fast enough, markups converge to constants fast enough that this correlation disappears.

Conditional Moment Restrictions and Convolution Equation.

Theorem L.2. *Suppose that Assumptions 1-3, regularity condition 1, and conditions of Theorem L.1 hold. Then, for any* $x_k \in \mathbf{D}_{x_k}$*,* $(\alpha', \beta'_k, \Gamma', F')$ *satisfies moment conditions* (D.1) *if and only if the following equation*

$$\int H(t; \alpha', \beta', \Gamma', F')\Lambda_H(t + \frac{\beta_k}{\alpha}x_k; f_{\xi,c})dt = 0, \tag{L.1}$$

holds, where

$$\begin{aligned}
H(t; \alpha', \beta', \Gamma', F') &= s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, t; x^{(2)}); x^{(2)}, \Gamma, F); x^{(2)}, \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t; x^{(2)}) \\
&\quad + \left(\alpha' - \frac{\beta'_k}{\beta_k}\alpha\right)(\mu_c - t) - \left(\beta'_{-k} - \frac{\beta'_k}{\beta_k}\beta_{-k}\right)x_{-k}, \\
\Lambda_H(\lambda; f_{\xi,c}) &= \int \alpha f_{\xi,w}(\alpha(c - \lambda) - \beta_{-k}x_{-k} - \eta, c)dc,
\end{aligned}$$

where $f_{\xi,c}$ *is the density function of* (ξ, c) *and* μ_c *is the expectation of* c*.*

The proof is similar to that of Theorem D.1.
According to Condition 2 of Theorem L.1, β_k/α is identified. Then, we can further simplify H:

$$\begin{aligned}
H(t; \alpha', \beta', \Gamma', F') &= s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p_{\mathbf{J}}(0, t; x^{(2)}); x^{(2)}, \Gamma, F); x^{(2)}, \Gamma', F') + \alpha' p_{\mathbf{J}}(0, t; x^{(2)}) \\
&\quad - \left(\beta'_{-k} - \frac{\beta'_k}{\beta_k}\beta_{-k}\right)x_{-k}.
\end{aligned}$$

Unique Solution for Convolution Equation (L.1) **and Identification of Product-Level Market Share Functions.** Define

$$\Theta = \{(\alpha', \beta', \Gamma', F') : \text{(D.1) hold at } (\alpha', \beta', \Gamma', F') \text{ for } x_k \in \mathbf{D}_{x_k}\},$$

and

$$\Theta_H = \{(\alpha', \beta', \Gamma', F') : \text{(L.1) holds for } x_k \in \mathbf{D}_{x_k}\}$$

Theorem L.2 establishes $\Theta = \Theta_H$. Define $\Theta_H^0 = \{(\alpha', \beta', \Gamma', F') : H(\cdot; \alpha', \beta', \Gamma', F') = 0\}$. Note that the true parameters $(\alpha, \beta, \Gamma, F) \in \Theta_H^0 \subset \Theta_H = \Theta$. Then, a necessary condition for the identification of $(\alpha, \beta, \Gamma, F)$ by moment conditions (D.1), i.e. $\Theta = \{(\alpha, \beta, \Gamma, F)\}$, is $\Theta_H^0 = \Theta_H$, i.e. $H = 0$ is the unique solution for convolution equation (L.1). This is the completeness of the location families generated by $\Lambda_H(\cdot; f_{\xi,c})$. Similar to Theorem D.2, the next theorem characterises the implications of this completeness:

Theorem L.3. *Suppose that conditions of Theorem L.2 hold.*

1. *If* $\Theta = \{(\alpha, \beta, \Gamma, F)\}$*, then* $\Theta_H^0 = \Theta_H$*.*

2. *If* $\Theta_H^0 = \Theta_H$*, then* α *and* $s_{\mathbf{J}\cdot}(\cdot; x_{t\mathbf{J}}^{(2)}, \Gamma, F)$ *are identified.*

The proof is similar to that of Theorem D.2.

Sufficient Conditions for the Completeness of Location Families. As in Theorem D.3, we propose the following result:

Theorem L.4. *Suppose that* (L.1) *holds for* $x_k \in \mathbb{R}^J$*.*

- *If* $f_{\xi,c}$ *is Gaussian, then* $\Theta_H^0 = \Theta_H$*.*
- *If the DGP is a model of demand for multiple products across categories, then under regularity conditions 2,* $\Theta_H^0 = \Theta_H$*.*

Note that H can be written as

$$H(p; \alpha', \beta', \Gamma', F') = s_{\mathbf{J}\cdot}^{-1}(s_{\mathbf{J}\cdot}(-\alpha p; x^{(2)}, \Gamma, F); x^{(2)}, \Gamma', F') + \alpha' p - \left(\beta'_{-k} - \frac{\beta'_k}{\beta_k}\beta_{-k}\right) x_{-k}.$$

The proofs of both statements are then similar to those of Theorem D.3. Finally, combining Theorems L.2-L.4, I obtain Theorem L.1.