An Empirical Model of Quantity Discounts with Large Choice Sets

Alessandro Iaria and Ao Wang

October 2021 No: 1378

Warwick Economics Research Papers

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

An Empirical Model of Quantity Discounts with Large Choice Sets

Alessandro Iaria and Ao Wang*

October 20, 2021

Abstract

We introduce a Generalized Nested Logit model of demand for bundles that can be estimated sequentially and virtually eliminates any challenge of dimensionality related to large choice sets. We use it to investigate quantity discounts for carbonated soft drinks by simulating a counterfactual with linear pricing. The prices of quantities up to 1L decrease by −31.5% while those of larger quantities increase by +14.8%. Purchased quantities decrease by −20.4%, associated added sugar by −23.8%, and industry profit by −20.5%. Consumer surplus however reduces only moderately, suggesting that linear pricing may be effective in limiting added sugar intake.

Keywords: Quantity Discounts, Large Choice Sets, Demand for Bundles, Generalized Nested Logit, Carbonated Soft Drinks, Purchase of Multiple Units.
JEL Codes: C55, C63, L4, L13, L66.

1 Introduction

In many important markets, firms offer nonlinear price schedules in which unit-prices vary with product size or quality.[1] Quantity discounts represent a common form of nonlinear pricing, where firms offer lower unit-prices for purchases of larger quantities. They enable firms to increase profits by screening between high-quantity and low-quantity consumers but can be detrimental for some groups of consumers (Crawford and Shum, 2007; Maskin and Riley, 1984; McManus, 2007; Mussa and Rosen, 1978). Despite their widespread diffusion in everyday life (e.g., packaged goods, mobile internet, and newspapers) and in services (e.g., mobile internet and newspapers) and a vast theoretical literature, there are relatively few empirical studies of quantity discounts.[2]

*University of Bristol and CEPR (alessandro.iaria@bristol.ac.uk) and University of Warwick (ao.wang@warwick.ac.uk). We would like to thank IRI for making the data available. All estimates and analysis in this paper based on data provided by IRI are by the authors and not by IRI.

[1] In contrast, recent findings suggest that many US retail chains may not take full advantage of other forms of price discrimination and tend to charge uniform prices for any given quantity of a product or service across different locations (Adams and Williams, 2019; Cho and Rust, 2010; DellaVigna and Gentzkow, 2019).

[2] See Anderson and Renault (2011) and Armstrong (2016) for a summary of the theoretical literature and below for an overview of the empirical studies.

This is partly motivated by the practical complexity of demand estimation in the context of bundles or multiple units, which usually involves large choice sets. As is well known, the estimation of demand for bundles is subject to a challenge of dimensionality in the number of products: the number of ways in which consumers can combine products into bundles can grow steeply in the number of products and the number of parameters capturing unobserved synergies among products within bundles can quickly become too large to be handled numerically (Berry et al., 2014). As a result, empirical papers estimating demand for bundles typically focus on applications with restricted choice sets, e.g. three products in Gentzkow (2007), or make restrictive assumptions on the form of unobserved preference heterogeneity, e.g. a multinomial logit in Ruiz et al. (2020).

We tackle this challenge and propose novel methods to estimate demand for bundles in the presence of large choice sets. We propose a Generalized Nested Logit (GNL) model, called Product-Overlap Nested Logit (PONL), that has as many overlapping nests as products and where each bundle belongs to all the nests corresponding to its product components (the standard nested logit being only a special case). Because of the overlapping nests, the PONL model cannot be estimated on the basis of Berry (1994) and, because of the large choice sets, Berry et al. (1995) may be impractical. We instead devise an optimization- and derivative-free iterative procedure that can be parallelized over both bundles and markets, virtually eliminating any challenge of dimensionality due to large choice sets.

As first argued by Gentzkow (2007), not accounting for correlation in the unobserved preferences of different products may confound the identification of complementarity and substitutability (Allen and Rehbeck, 2019, 2020; Ershov et al., 2021; Fox and Lazzati, 2017; Iaria and Wang, 2019, 2021; Wang, 2019), stressing the importance of allowing for flexible forms of unobserved heterogeneity in the specification of demand for bundles. In applications with thousands of bundles (like the one we study), the estimation of nonparametric models (Compiani, 2019) or even just of mixed logit models (Gentzkow, 2007; Iaria and Wang, 2019; Liu et al., 2010) may however be prohibitive. As a practical alternative, we propose the PONL model in which there is a nest for each product and every bundle belongs to as many nests as the different products it includes. In a standard nested logit model where each bundle only belongs to one nest, either all bundles have equally correlated preferences (a unique nest) or some of the bundles with overlapping products have uncorrelated preferences (more than one nest, Song et al., 2017). Differently, the PONL model allows for correlation in the unobserved preferences to depend on the degree of overlap in the composition of products between any two bundles.

An essential factor behind the practical advantages of the proposed estimator is the use of individual-level purchases in the aggregate form of bundle-level purchase probabilities. As shown by Berry (1994), working with purchase probabilities sometimes allows one to re-write complex non-linear demand models as linear regressions that are easier to estimate. Because of the overlapping nests, Berry (1994)'s classic 2SLS regression does not apply to the PONL model.[3] Given observations on bundle-level purchase probabilities, we however show that the PONL model can be estimated by

[3] As will be clear below, the presence of overlapping nests implies a lack of observability of the within-nest purchase probabilities, which are typically used as explanatory variables in Berry (1994)'s regression.

a *constrained* 2SLS and implemented by a convenient iterative procedure. Importantly, the proposed estimator is robust to price endogeneity and is easy to implement with large choice sets. Differently, both the classic approach by Berry et al. (1995) and its MPEC counterpart (Dubé et al., 2012; Su and Judd, 2012) would be impractical with large choice sets, mainly because of the large dimensionality of the demand system.[4] Also traditional likelihood-type estimators based on the direct use of individual-level purchases would not be computationally convenient with large choice sets, mainly because of the large number of fixed effects required to control for price endogeneity (Iaria and Wang, 2019).

In extensive Monte Carlo simulations, we investigate the numerical and finite sample properties of the proposed iterative procedure for the estimation of the PONL model. We show that our estimator can be successfully implemented on standard computers with choice sets of approximately $20,000$ bundles and around $4,000,000$ demand parameters. The proposed iterative procedure allows for a complete parallelization across bundles and markets. In this sense, its numerical convenience increases in the number of CPU cores, and is therefore expected to improve over time as these become more cheaply available. We illustrate that, in addition to being numerically convenient, the proposed iterative procedure has desirable finite sample properties and delivers precise estimates even with large choice sets and correspondingly large numbers of demand parameters.

We implement our methods to investigate the welfare consequences of quantity discounts in the market for carbonated soft drinks (CSDs) in the USA. Using household-level purchase data by IRI for the period 2008-2011 (Bronnenberg et al., 2008), we document that households commonly purchase multiple units of CSDs on any shopping trip (6.6L on average) (Chan, 2006; Dubé, 2004; Ershov et al., 2021) and the pervasiveness of quantity discounts for purchases involving larger quantities (e.g., the average unit-price of a Diet Coke is higher for a 12oz can than for a 2L bottle).[5]

We observe that, according to intuition, larger households tend to purchase larger quantities of CSDs, both as multiple units of the same product and as combinations of different products. Despite being unable to price discriminate directly on the basis of household size (third-degree price discrimination), firms may rely on quantity discounts as a screening device to induce households of different sizes to self-select alternative prices (Maskin and Riley, 1984; Mussa and Rosen, 1978). We however document that, because also single-person households purchase multiple units of CSDs, quantity discounts only achieve imperfect screening among households of different sizes. In this complex situation of imperfect screening in an oligopolistic market with differentiated products, the welfare effects of quantity discounts are ambiguous (Anderson and Leruth, 1993; Armstrong, 2013; Varian, 1989).

We then estimate a flexible PONL model with around $16,900$ bundles of CSDs and $176,700$ demand parameters and empirically assess the welfare effects of the observed quantity discounts by simulating a counterfactual with linear pricing (i.e., forcing constant unit-prices for all products).

[4] For example, in our application we observe around $16,900$ bundles purchased over $1,200$ markets. In the classic fixed point implementation of Berry et al. (1995), this would require the computation of $1,200$ separate demand inverses with up to $16,900$ equations each. Analogously, the MPEC implementation of Berry et al. (1995) would require the computation of up to $1,200 \times 16,900$ nonlinear constraints and their derivatives.

[5] O'Connell and Smith (2020) document similar patterns for the market of soft drinks in the UK.

Our counterfactual simulations suggest that linear pricing would lead to a reduction of −31.5% in the average price of small quantities (up to one liter) and to an increase of +14.8% in the average price of larger quantities (more than one liter), making purchases of smaller quantities relatively more attractive for all households. While such drastic price changes would have important consequences on quantity purchased and industry profit, they would have less of an impact on consumer surplus.

Total quantity purchased would decrease by −20.4% and, as a consequence, industry profit would shrink by −20.5%. Despite the substantial reduction in quantity purchased, consumer surplus would not however reduce too sharply, with a compensating variation of +3.6$ per household-year (amounting to 2.7% of total expenditure on CSDs). This is the result of two intuitive countervailing forces: on the one hand, consumer surplus would decrease because of the contraction in purchases of larger quantities at relatively higher prices; on the other, consumer surplus would increase because of the more frequent purchases of single units at relatively lower prices. While the negative effect would slightly dominate the positive for all households, there would still be some heterogeneity: multi-person households would substitute less away from the more expensive larger quantities toward the cheaper small quantities, facing larger losses in consumer surplus (a compensating variation of +3.9$ as opposed to +1.5$).

These results open up an important avenue for future research: a ban on quantity discounts could serve as a previously unexplored policy tool to limiting the consumption of CSDs and the intake of added sugar (Allcott et al., 2019; Bollinger et al., 2011; Dubois et al., 2020; O'Connell and Smith, 2020; Wang, 2015). Ricciuto et al. (2021) report that in the USA, in the period 2011-2012, 42.4% of the added sugar intake came from CSDs. Linear pricing would lead households to drastically reduce the purchased quantities of CSDs while only marginally reducing consumer surplus, potentially causing large reductions in added sugar intake at the expense of a contraction in industry profit but none of the extra information (e.g., quantifying the marginal externality of added sugar) required to implement effective sugar taxes (Allcott et al., 2019; O'Connell and Smith, 2020). Our back-of-the-envelope calculations suggest that linear pricing would indeed reduce added sugar intake from CSDs by −23.8%, a similar amount as that implied by various sugar taxes in the UK and USA.[6]

There is a large empirical literature leveraging the estimation of demand for bundles.[7] Part of this literature investigates quantity discounts, as for example: Allenby et al. (2004); Aryal and Gabrielli (2020); Crawford and Shum (2007); Ivaldi and Martimort (1994); Leslie (2004); Levitt et al. (2016); Liu et al. (2010); Luo (2018); McManus (2007); McManus et al. (2020); Shiller and Waldfogel (2011). Because of the challenge of dimensionality in the number of products, papers in this empirical literature

[6] O'Connell and Smith (2020) find that an optimal sugar tax in the UK would result in a decrease of −28.4% in the purchased quantities of added sugar from soft drinks. Similarly, Dubois et al. (2020) find that a sugar tax of the form and size typically implemented in the UK and many US locations would lead to a reduction of around −21% in the purchased quantities of added sugar from soft drinks on-the-go. Seiler et al. (2021) document that a sugar tax introduced in Philadelphia led to a decrease of −16% in the purchased quantities of added sugar from soft drinks. There are also studies that do not find significant effects of sugar taxes in the USA on the reduction of purchased quantities of added sugar from soft drinks, such as Bollinger and Sexton (2018); Rojas and Wang (2017); Wang (2015).

[7] Some examples are: Crawford and Yurukoglu (2012); Florez-Acosta and Herrera-Araujo (2020); Fosgerau et al. (2021); Gentzkow (2007); Gentzkow et al. (2014); Hendel (1999); Ho et al. (2012); Manski and Sherman (1980); Ruiz et al. (2020); Thomassen et al. (2017).

either focus on applications with restricted choice sets or limited forms of unobserved heterogeneity. Our methods can enable empirical researchers to scale up the numbers of bundles while allowing for realistic forms of unobserved heterogeneity: demand across multiple product categories involving many products, like grocery or online shopping (Reimers and Waldfogel, 2021; Smith, 2006), mergers in markets with both substitutes and complements (Cournot, 1838; Ershov et al., 2021), mixed bundling pricing strategies (Adams and Yellen, 1976; Chu et al., 2011), (unintended) spillovers of taxes from a product category to others (Allcott et al., 2019; Dubois et al., 2020), and many more.

Two novel approaches to addressing large choice sets in the estimation of demand for bundles were recently proposed by Ershov et al. (2021) and Lewbel and Nesheim (2019). Ershov et al. (2021) allow for a very large number of products, but restrict the way they can be combined into bundles (at most two different products, one unit each) and the number of parameters capturing unobserved synergies among products (one per market, the same across all bundles). While this approach is appealing in applications with "many but small" bundles, ours is better suited to handle larger bundles involving multiple units of the same or of different products, such as in the case of quantity discounts. Lewbel and Nesheim (2019) instead depart from the use of standard discrete choice models and specify a more general discrete-continuous choice model along the lines of Dubin and McFadden (1984). They address the problem of large choice sets with sparsity, by assuming that each consumer tends to purchase positive quantities of only a few products. While allowing for more flexible unobserved heterogeneity than the PONL model, Lewbel and Nesheim (2019) assume prices to be exogenous. Our approach complements the one by Lewbel and Nesheim (2019) and is more suitable to applications in which consumers purchase larger varieties of products and/or price endogeneity is a concern.

The proposed PONL model contributes to the literature on Generalized Extreme Value (GEV) models started by McFadden (1978), which famously includes the multinomial logit and the nested logit. The PONL model is a specialization of Generalized Nested Logit (GNL) model (Abbe et al., 2007; Ben-Akiva and Bierlaire, 1999; Bierlaire, 2006; Wen and Koppelman, 2001), the most general example of GEV model, to the case of demand for bundles. Our econometric treatment of the PONL model contributes to a small but growing literature on the use of GEV models for the convenient estimation of demand from aggregate market-level data (Bresnahan et al., 1997; Davis and Schiraldi, 2014; Fosgerau et al., 2021; Grigolon, 2021), as initiated by Berry (1994) for the multinomial logit and the nested logit. In particular, our constrained 2SLS estimator and iterative procedure can be helpful also for the estimation of other GNL models with large choice sets.

Our empirical analysis contributes to the applied literature on the industry for CSDs. Some of the papers in this literature rely on the estimation of demand for multiple units (Chan, 2006; Dubé, 2004; Ershov et al., 2021; Hendel and Nevo, 2013; Wang, 2015), while many others direct their efforts to other important aspects of the industry, such as vertical relations between manufacturers and retailers or sugar taxes (Allcott et al., 2019; Bonnet and Dubois, 2010; Dubois et al., 2020; Huang and Liu, 2017; Molina, 2020; O'Connell and Smith, 2020). To the best of our knowledge, we are the first to explicitly investigate the welfare effects of quantity discounts in this industry.[8]

2 The Product-Overlap Nested Logit (PONL) Model

Let there be T independent markets indexed by $t \in \mathbf{T}$ and J products indexed by $j \in \mathbf{J}$ that can be purchased in isolation or in combination in each market. A bundle is any combination of products and number of units of each product (e.g., three units of j, one unit of k, and two units of r). Denote the set of single units of any product and (multi-unit) bundles by $\mathbf{C}_1$ and its size by $|\mathbf{C}_1| = C_1$, the full choice set by $\mathbf{C} = \mathbf{C}_1 \cup \{0\}$ and its size by $|\mathbf{C}| = C$, where 0 is the outside option of not purchasing anything. Denote the set of (multi-unit) bundles by $\mathbf{C}_2 = \mathbf{C}_1 \setminus \mathbf{J}$ and its size by $|\mathbf{C}_2| = C_2 = C_1 - J$. Each element of this set is a bundle made of multiple units of one or of different products.

As first argued by Gentzkow (2007), accounting for correlation in the unobserved preferences of different products is crucial for the identification of demand for bundles (Allen and Rehbeck, 2019, 2020; Ershov et al., 2021; Fox and Lazzati, 2017; Iaria and Wang, 2021; Wang, 2019). Each $\mathbf{b} \in \mathbf{C}_1$ is a combination products, and any pair of bundles will have a certain degree of overlap in terms of product components. It is then important to account for such overlapping structure and the potential correlation patterns this may imply among the unobserved preferences of different bundles. For example, the unobserved preferences of bundle (j, k) may differentially correlate to those of any other bundle that either includes only j (correlation only via j), only k (correlation only via k), both (correlation via both channels), or neither (lack of correlation).

On the one hand, simple models like the Multinomial Logit (MNL) or the Nested Logit (NL), which can be easily estimated with large choice sets (Crawford et al., forthcoming), cannot appropriately capture these intuitive patterns of correlation.[9] On the other hand, more appropriate non-parametric (Compiani, 2019) or even mixed logit models can be unfeasible in application with large choice sets (Gentzkow, 2007; Iaria and Wang, 2019; Liu et al., 2010).[10] As a solution, we propose a special case of Generalized Nested Logit (GNL) model (Abbe et al., 2007; Ben-Akiva and Bierlaire, 1999; Bierlaire, 2006; Wen and Koppelman, 2001) with overlapping nests that specifically accounts for the product overlap between bundles in terms of unobserved preferences but is still practical with large choice sets. We call this the Product-Overlap Nested Logit (PONL) model.

[8] Bonnet and Dubois (2010) do not study nonlinear pricing with respect to "final" consumers (as we do in this paper), but rather two-part tariff contracts between manufacturers and retailers. See also Bonnet et al. (2013) for a related analysis of the market for coffee. While close in spirit to our paper, Hendel and Nevo (2013) studies the welfare effects of intertemporal price discrimination (i.e., temporary price reductions) rather than quantity discounts.

[9] The MNL implies that the unobserved preferences of any two bundles are independent. The NL instead requires every bundle to belong uniquely to one nest, so that either all bundles have equally correlated preferences (a unique nest) or some of the bundles with overlapping components end up with uncorrelated preferences (more than one nest, Song et al., 2017). We return to this point below.

[10] See also discussion in section 4.3.

2.1 Unobserved Preferences and their Correlations Across Bundles

In the PONL model, each nest $\mathbf{N}_j$ is defined as the set of bundles that include at least one unit of product j, for $j = 1, ..., J$: $\mathbf{N}_j = \{\mathbf{b} \in \mathbf{C}_1 : \mathbf{b}$ includes at least one unit of $j\}$, while the outside option belongs to its own singleton nest $\mathbf{N}_0$. By construction, $\mathbf{N}_j$ and $\mathbf{N}_{j'}$ are overlapping as long as there exists at least a bundle $\mathbf{b}$ that includes both one unit of j and one of j'. The membership of $\mathbf{b}$ to nest $\mathbf{N}_j$ is determined by the allocation parameter $\omega_{\mathbf{b}j} = \mathbf{1}_{\mathbf{b}\in\mathbf{N}_j} \times (\sum_{j'=1}^{J} \mathbf{1}_{\mathbf{b}\in\mathbf{N}_{j'}})^{-1}$, where $\mathbf{1}_E$ denotes the indicator function for event E. Every $\omega_{\mathbf{b}j}$ is observed and equals either zero if $\mathbf{b} \notin \mathbf{N}_j$ ($\mathbf{b}$ does not include any unit of j) or one divided by the number of nests $\mathbf{b}$ belongs to, if $\mathbf{b} \in \mathbf{N}_j$.

We derive the PONL model as a combination of NL models, following the representation first proposed by Abbe et al. (2007) for the GNL model. While this representation is not necessary to derive the PONL model, it clarifies the connection with the NL and the ways in which the PONL generalizes it.[11] Denote by $U_{it\mathbf{b}j}$ the NL indirect utility of household i in market t from purchasing $\mathbf{b}$, as if $\mathbf{b}$ belonged uniquely to nest $\mathbf{N}_j$:

$$U_{it\mathbf{b}j} = \delta_{t\mathbf{b}} + \eta_{itj} + \lambda_j \varepsilon_{it\mathbf{b}j}, \tag{1}$$

where $\delta_{t\mathbf{b}}$ is the average utility of $\mathbf{b}$ among the households in market t, $\eta_{itj} + \lambda_j \varepsilon_{it\mathbf{b}j}$ is the usual unobserved component of preferences that gives rise to the NL model (Berry, 1994; Cardell, 1997), η_{itj} is common to all bundles in nest $\mathbf{N}_j$ and introduces correlation in their unobserved preferences, and $\lambda_j \in (0, 1]$ is the nesting parameter that determines the strength of such correlation. Given (1), the PONL indirect utility of household i in market t from purchasing $\mathbf{b}$ can then be expressed as:[12]

$$\begin{aligned} U_{it\mathbf{b}} &= \max_{j\in\mathbf{J}} \{\ln \omega_{\mathbf{b}j} + U_{it\mathbf{b}j}\} \\ &= \max_{j\in\mathbf{J}} \{\delta_{t\mathbf{b}} + \ln \omega_{\mathbf{b}j} + \eta_{itj} + \lambda_j \varepsilon_{it\mathbf{b}j}\}, \end{aligned} \tag{2}$$

where $\ln \omega_{\mathbf{b}j} + \eta_{itj} + \lambda_j \varepsilon_{it\mathbf{b}j} = -\infty$ for any j corresponding to a nest that does not include $\mathbf{b}$ (given that $\omega_{\mathbf{b}j} = 0$). Then, the elements of $(\ln \omega_{\mathbf{b}j} + \eta_{itj} + \lambda_j \varepsilon_{it\mathbf{b}j})_{j=1}^{J}$ that contribute to the maximum in (2) are those corresponding to the nests that include $\mathbf{b}$. This implies that (2) simplifies to (1), i.e. the indirect utility of the NL model, for any $\mathbf{b}$ that only belongs to one nest, say j, given that $\max_{j\in\mathbf{J}} \{\ln \omega_{\mathbf{b}j} + U_{it\mathbf{b}j}\} = U_{it\mathbf{b}j}$ in such case. Differently, for any $\mathbf{b}$ that belongs to multiple nests, $U_{it\mathbf{b}}$ is determined by the *largest* NL indirect utility $U_{it\mathbf{b}j}$ among the nests that include $\mathbf{b}$.

The PONL indirect utility (2) then implies the following correlation structure among the unob-

[11]Alternatively, one can derive the PONL model by an appropriate choice of generating function, as originally proposed by McFadden (1978) for any Generalized Extreme Value model. For this more direct but less economically intuitive derivation of the GNL model, see Wen and Koppelman (2001).

[12]Bierlaire (2006) shows that any GNL model, and hence also the PONL model, is consistent with random utility maximization when $\lambda_j \in (0, 1]$, $j = 1, ..., J$ (McFadden, 1978).

served preferences of different bundles:[13]

$$\mathrm{Corr}(U_{it\mathbf{b}}, U_{it\mathbf{b}'}) = \mathrm{Corr}\left(\max_{j\in\mathbf{J}}\left\{\ln\omega_{\mathbf{b}j} + \eta_{itj} + \lambda_j\varepsilon_{it\mathbf{b}j}\right\}, \max_{j\in\mathbf{J}}\left\{\ln\omega_{\mathbf{b}'j} + \eta_{itj} + \lambda_j\varepsilon_{it\mathbf{b}'j}\right\}\right), \tag{3}$$

with

$$\mathrm{Corr}_j(\eta_{itj} + \lambda_j\varepsilon_{it\mathbf{b}j}, \eta_{itk} + \lambda_k\varepsilon_{it\mathbf{b}'k}) = \begin{cases} 0 & k \neq j \\ \left(1-\lambda_j^2\right)\mathbf{1}_{\mathbf{b},\mathbf{b}'\in\mathbf{N}_j} & k = j. \end{cases} \tag{4}$$

This highlights that any pair of elements from the maxima in (3) has correlation corresponding to that of the NL model in (4). In fact, note that (3) simplifies to (4) in the case of the NL model. Starting from the PONL model, the NL model can be obtained by setting, for each $\mathbf{b} \in \mathbf{C}$, $\omega_{\mathbf{b}j} = 1$ for any one nest j and $\omega_{\mathbf{b}j'} = 0$ for every other nest $j' \neq j$. Suppose that bundle $\mathbf{b}$ belongs to nest j. In the NL model, (3) then implies that $\mathrm{Corr}(U_{it\mathbf{b}}, U_{it\mathbf{b}'}) = 1 - \lambda_j^2$ if also $\mathbf{b}'$ belongs to nest j, or zero otherwise. Differently, in the PONL model, $\mathrm{Corr}(U_{it\mathbf{b}}, U_{it\mathbf{b}'})$ will be a function of all the nesting parameters λ_j, $j = 1, ..., J$, corresponding to the nests $\mathbf{N}_j$, $j = 1, ..., J$, that include both bundles $\mathbf{b}$ and $\mathbf{b}'$.

Example 1. The possibility of any bundle to belong to multiple nests plays an important conceptual role in empirical models of demand for bundles: for each bundle is a combination of products, any product will typically be part of several bundles. Without overlapping nests, the unobserved preferences of any two bundles from different nests will be uncorrelated. To see why this can be unrealistic, suppose there are three products 1, 2, and 3 and that households can buy them in isolation or can jointly buy 1 and 2, so that the choice set is $\mathbf{C} = \mathbf{C}_1 \cup \{0\}$ with $\mathbf{C}_1 = \{1, 2, 3, (1, 1), (2, 2), (1, 2)\}$.

The NL model would require to uniquely and arbitrarily allocate each element of $\mathbf{C}_1$ to a nest. For example, one could specify three nests: $\mathbf{N}_i = \{1, (1, 1), (1, 2)\}$, $\mathbf{N}_{ii} = \{2, (2, 2)\}$, and $\mathbf{N}_{iii} = \{3\}$. This, however, is not fully satisfactory. While it is true that the alternatives *within* each nest share some common feature, i.e. product 1 in $\mathbf{N}_i$, product 2 in $\mathbf{N}_{ii}$, and product 3 in $\mathbf{N}_{iii}$, it would be desirable that also bundle $(1, 2)$ shared common features with the elements of *both* $\mathbf{N}_i$ and $\mathbf{N}_{ii}$. In general, the NL model cannot accommodate this intuitive requirement for all products and bundles: because bundle $(1, 2)$ can only be allocated to either $\mathbf{N}_i$ or $\mathbf{N}_{ii}$, its unobserved preferences will either have correlation $1-\lambda_i^2$ with those of $\{1, (1, 1)\}$ or $1-\lambda_{ii}^2$ with those of $\{2, (2, 2)\}$, but will not correlate with both.[14] Any nesting structure in the NL must partition $\mathbf{C}$, ruling out correlation among at least some of the bundles with overlapping components (Song et al., 2017).

The PONL model addresses this limitation in a convenient way and without requiring arbitrary specifications of the nests. Each product and bundle is automatically allocated to one or more of $J = 3$ nests: $\mathbf{N}_1 = \{1, (1, 1), (1, 2)\}$, $\mathbf{N}_2 = \{2, (2, 2), (1, 2)\}$, and $\mathbf{N}_3 = \{3\}$. Any $\mathbf{b}$ that uniquely belongs to

[13] A closed-form expression for $\mathrm{Corr}(U_{it\mathbf{b}}, U_{it\mathbf{b}'})$ has not yet been derived for the GNL model. The expression we present in (3) was first derived by Abbe et al. (2007) and is the most interpretable characterization we are aware of. To improve understanding, some authors have also proposed interesting approximations (Marzano et al., 2013) and simulations (Marzano and Papola, 2008).

[14] Specifying instead $\mathbf{C}_1$ as a unique nest would rule out the possibility of 1 being more closely related to $(1, 1)$ and $(1, 2)$ than to 3, and similarly for 2.

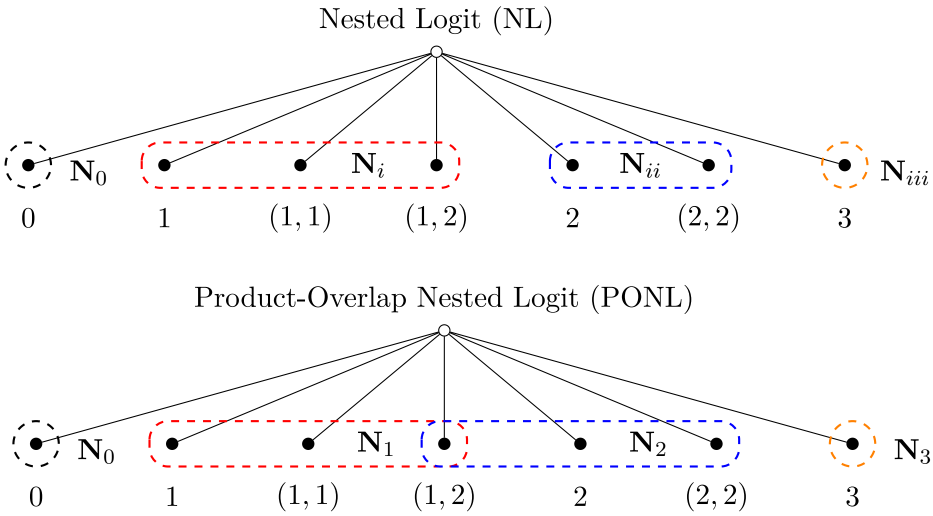


Figure 1: Nesting Structures of NL and PONL

nest $\mathbf{N}_j$ has allocation parameters $\omega_{\mathbf{b}j} = 1$ and $\omega_{\mathbf{b}j'} = 0$ for $j' \neq j$, so that: $\omega_{11} = \omega_{(1,1)1} = \omega_{22} = \omega_{(2,2)2} = \omega_3 = 1$ and $\omega_{12} = \omega_{13} = \omega_{21} = \omega_{23} = \omega_{31} = \omega_{32} = \omega_{(1,1)2} = \omega_{(1,1)3} = \omega_{(2,2)1} = \omega_{(2,2)3} = 0$. Moreover, $(1,2)$, which belongs to multiple nests, has allocation parameters: $\omega_{(1,2)1} = \omega_{(1,2)2} = 0.5$ and $\omega_{(1,2)3} = 0$. Figure 1 visualizes the nesting structures of the NL and the PONL.

In the PONL model, the unobserved preferences of bundle $(1,2)$ will be allowed to correlate both with those of the bundles in $\mathbf{N}_1$ (that include at least one unit of product 1) and with those of bundles in $\mathbf{N}_2$ (that include at least one unit of product 2), and potentially to different degrees on the basis of λ_1 and λ_2. We find this intuitively appealing in the context of demand for bundles, in that the PONL model naturally accommodates correlation in the unobserved preferences among bundles on the basis of their degree of overlap in the composition of products. ■

2.2 Average Utilities and Demand Synergies

With some abuse of notation, we refer to the components of a bundle $\mathbf{b}$ simply as "products" and denote them by $j \in \mathbf{b}$. Despite this shortcut, we stress that bundles can contain multiple units of a single product. In addition, we maintain throughout that for any bundle $\mathbf{b} \in \mathbf{C}_2$, a single unit of each product $j \in \mathbf{b}$ can also be purchased in isolation. This rules out the complication that some product can only be purchased through bundles.[15] We denote by δ_{tj} the market t-specific average utility of a single unit of product j and, as is common in applied work, we assume it to be linear:

$$\delta_{tj} = \delta_j + x_{tj}\beta - \alpha p_{tj} + \xi_{tj}, \tag{5}$$

[15] In Appendix E, we discuss a simple procedure to extend the use of our estimator to applications in which some product can only be purchased through bundles. We implement this procedure in the empirical analysis in section 6.

where δ_j is an intercept, x_{tj} is a K-dimensional vector of characteristics, p_{tj} is the price of a single unit of product j in market t, (β, α) are preference parameters, and ξ_{tj} is a residual observed by all economic agents (e.g., households and firms) but unobserved by the econometrician. We assume that the $K \times J$ characteristics are exogenous in each market t:

$$\mathbb{E}\left[(\xi_{tj})_{j=1}^{J} \middle| (x_{tj})_{j=1}^{J}\right] = 0. \tag{6}$$

Differently, the prices $(p_{tj})_{j=1}^{J}$ could be set by firms on the basis of $(\xi_{tj})_{j=1}^{J}$ and therefore correlate with these unobservables. Following Gentzkow (2007), we denote by $\delta_{t\mathbf{b}} = \sum_{j\in\mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$ the market t-specific average utility of bundle $\mathbf{b} \in \mathbf{C}_2$. For example, if $\mathbf{b} = (j, j, k)$, i.e. two units of product j and one of product k, then $\delta_{t(j,j,k)} = 2\delta_{tj} + \delta_{tk} + \Gamma_{t(j,j,k)}$. We refer to $\Gamma_{t\mathbf{b}}$ as the demand synergy parameter, which captures the extra average utility from purchasing the products in bundle $\mathbf{b}$ jointly rather than separately. In Gentzkow (2007)'s demand for on-line and printed newspapers, $\Gamma_{t\mathbf{b}}$ represents synergies in the consumption of different news outlets. However, demand synergies can also arise for other reasons, such as shopping costs (Florez-Acosta and Herrera-Araujo, 2020; Pozzi, 2012; Thomassen et al., 2017) or aggregation across multiple choices (Dubé, 2004; Hendel, 1999).[16] In the context of quantity discounts, for example, even excluding any other source of synergies, $\Gamma_{t\mathbf{b}} = -\alpha(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}) > 0$ whenever it is cheaper to purchase the products in bundle $\mathbf{b}$ jointly rather than separately, i.e. $p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj} < 0$.

Throughout the presentation of the model and estimator, we remain agnostic about the market t-specific demand synergies $\Gamma_t = (\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$, and treat them as parameters to be estimated. In applications with observable bundle-level characteristics, one can however project these parameters onto observables and learn more about their nature (as we do in our application with quantity discounts).

2.3 Purchase Probabilities

Denote by $\mathscr{s}_{t\mathbf{b}}$, $\mathscr{s}_{t(\mathbf{b}|j)}$, $\mathscr{s}_t^j$, and $\mathscr{s}_{t0}$ the t-specific purchase probabilities of, respectively: $\mathbf{b}$ unconditional on any nest, $\mathbf{b}$ conditional on nest $\mathbf{N}_j$, any bundle in nest $\mathbf{N}_j$, and the outside option. Similar to the NL model, also in the PONL model any $\mathbf{b}$ that uniquely belongs to nest j has purchase probability $\mathscr{s}_{t\mathbf{b}} = \mathscr{s}_{t(\mathbf{b}|j)}\mathscr{s}_t^j$. Any $\mathbf{b}$ that instead belongs to multiple nests has $\mathscr{s}_{t\mathbf{b}}$ given by the sum of the joint purchase probabilities $\mathscr{s}_{t(\mathbf{b}|j)}\mathscr{s}_t^j$ over the $J + 1$ nests, where $\mathscr{s}_{t(\mathbf{b}|k)} = 0$ for any k such that $\mathbf{b} \notin \mathbf{N}_k$.

Given (1), (2), and denoting $V_{it\mathbf{b}j} = \delta_{t\mathbf{b}} + \ln \omega_{\mathbf{b}j} + \lambda_j \varepsilon_{it\mathbf{b}j}$, the purchase probability of $\mathbf{b} \in \mathbf{C}_1$ in

[16] If households face shopping costs every time they visit a store, they may prefer to purchase all their products at once rather than over several trips (one-stop shoppers). Moreover, if households delegate grocery shopping to one person, then the need to accommodate different requests within the household may lead to the purchase of multiple units of the same or of different products on any shopping trip.

market t can be obtained as:[17]

$$\begin{aligned} \mathscr{s}_{t\mathbf{b}} &= \Pr(i \text{ in } t \text{ purchases } \mathbf{b}) \\ &= \sum_{j=0}^{J} \Pr(i \text{ in } t \text{ purchases } \mathbf{b} | i \text{ in } t \text{ purchases bundle in } \mathbf{N}_j) \Pr(i \text{ in } t \text{ purchases bundle in } \mathbf{N}_j) \\ &= \sum_{j=0}^{J} \Pr\left(V_{it\mathbf{b}j} > V_{it\mathbf{b}'j}, \forall \mathbf{b}' \neq \mathbf{b}, \mathbf{b}' \in \mathbf{N}_j\right) \Pr\left(\max_{\mathbf{b}' \in \mathbf{N}_j} \{V_{it\mathbf{b}'j} + \eta_{itj}\} > \max_{\mathbf{b}' \in \mathbf{N}_\ell} \{V_{it\mathbf{b}'\ell} + \eta_{it\ell}\}, \forall \ell \neq j\right) \\ &= \sum_{j=0}^{J} \underbrace{\frac{(\omega_{\mathbf{b}j} \exp(\delta_{t\mathbf{b}}))^{1/\lambda_j}}{\sum_{\mathbf{b}' \in \mathbf{N}_j} (\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j}}}_{\mathscr{s}_{t(\mathbf{b}|j)}} \underbrace{\frac{\left(\sum_{\mathbf{b}' \in \mathbf{N}_j} (\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j}\right)^{\lambda_j}}{\sum_{\ell=0}^{J} \left(\sum_{\mathbf{b}' \in \mathbf{N}_\ell} (\omega_{\mathbf{b}'\ell} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_\ell}\right)^{\lambda_\ell}}}_{\mathscr{s}_t^j}. \end{aligned} \tag{7}$$

Because the outside option belongs to its own singleton nest $\mathbf{N}_0$, by further assuming $\delta_{t0} = 0$ we get:[18]

$$\mathscr{s}_{t0} = \frac{1}{\sum_{\ell=0}^{J} \left(\sum_{\mathbf{b}' \in \mathbf{N}_\ell} (\omega_{\mathbf{b}'\ell} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_\ell}\right)^{\lambda_\ell}}. \tag{8}$$

2.4 Demand Inverse

Similar to the NL studied by Berry (1994), also the PONL purchase probabilities (7) and (8) can be conveniently "inverted" with respect to the average utilities, giving rise for any bundle $\mathbf{b} \in \mathbf{C}_2$ to:

$$\begin{aligned} \ln \mathscr{s}_{t\mathbf{b}} - \ln \mathscr{s}_{t0} &= \ln \left(\sum_{j=0}^{J} (\omega_{\mathbf{b}j} \exp(\delta_{t\mathbf{b}}))^{1/\lambda_j} \left(\sum_{\mathbf{b}' \in \mathbf{N}_j} (\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j} \right)^{\lambda_j - 1} \right) \\ &= \delta_{t\mathbf{b}} + \ln \left(\sum_{j=0}^{J} \omega_{\mathbf{b}j} \left(\mathscr{s}_{t(\mathbf{b}|j)} \right)^{1-\lambda_j} \right) \\ &= \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}} + \ln \left(\sum_{j \in \mathbf{b}} \omega_{\mathbf{b}j} \left(\mathscr{s}_{t(\mathbf{b}|j)} \right)^{1-\lambda_j} \right). \end{aligned} \tag{9}$$

Different from a NL model, however, the possibility of overlapping nests leads λ_j, $j \in \mathbf{b}$, to be nonlinear in (9). In unrestricted versions of the GNL model, analogous nonlinearities appear also in the equations corresponding to a single unit of any product. However, in the special case of the PONL model, any single unit of product j has allocation parameters $\omega_{jj} = 1$ and $\omega_{jj'} = 0$ for $j' \neq j$, so that:

$$\ln \mathscr{s}_{tj} - \ln \mathscr{s}_{t0} = \delta_{tj} + (1 - \lambda_j) \ln(\mathscr{s}_{t(j|j)}). \tag{10}$$

Plugging (5) into (10) and, respectively, into (9), we obtain:

$$\ln \mathscr{s}_{tj} - \ln \mathscr{s}_{t0} = \delta_j + x_{tj}\beta - \alpha p_{tj} + (1 - \lambda_j) \ln(\mathscr{s}_{t(j|j)}) + \xi_{tj}. \tag{11}$$

[17] See footnote 11.

[18] In the context of demand for bundles, this otherwise standard normalization has important repercussions for the identification of the demand synergy parameters. We discuss this in detail in the empirical application.

$$\Gamma_{t\mathbf{b}} = \ln(\mathscr{s}_{t\mathbf{b}}) - \ln(\mathscr{s}_{t0}) - \sum_{j \in \mathbf{b}} (\delta_j + x_{tj}\beta - \alpha p_{tj} + \xi_{tj}) - \ln \left(\sum_{j \in \mathbf{b}} \omega_{\mathbf{b}j} \left(\mathscr{s}_{t(\mathbf{b}|j)} \right)^{1-\lambda_j} \right). \tag{12}$$

3 Identification

An important and distinctive feature of our approach is to use different parts of system (9) to identify and estimate different parts of the PONL model. We restrict attention to the $T \times J$ linear equations in (11), corresponding to the purchases of *single units*, for the identification and estimation of $(\delta, \beta, \alpha, \lambda)$, and then rely on the remaining $T \times C_2$ nonlinear equations in (12), corresponding to the purchases of *multiple units*, for the identification and estimation of the demand synergies $(\Gamma_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$, $t \in \mathbf{T}$. Alternatively, one could identify and estimate the entire PONL model simultaneously from the $T \times C_1$ nonlinear equations in (9). While both approaches are possible, we pursue the former because the latter leads to a more complex problem of endogeneity (as discussed in the next section) and practically less convenient estimators, especially in applications with large choice sets (as discussed in section 4).

3.1 Endogeneity in System (9)

By relying *sequentially* on (11) and (12) for the identification and estimation of the PONL model, we face a practically simpler problem of endogeneity than by relying *simultaneously* on system (9). Intuitively, our approach only uses the $T \times J$ equations for single units in (11) as a linear regression to learn about $(\delta, \beta, \alpha, \lambda)$ and then uses the remaining $T \times C_2$ equations for multiple units in (12) as a plug-in to learn about the demand synergies. This way, the problem of endogeneity is limited to the linear regression in (11), i.e. the correlation of $(p_{tj}, \mathscr{s}_{t(j|j)})$ with ξ_{tj}, and can be addressed by instruments that satisfy moment conditions at the level of the single unit j.[19] Differently, the simultaneous use of all the equations in (9) would lead to a more complex problem of endogeneity that can only be addressed by moment conditions both at the single unit j and at the bundle $\mathbf{b}$ level.

To illustrate this additional complexity, suppose to observe prices and characteristics both of single units $(p_{tj}, x_{tj})_{j=1}^{J}$ and of bundles $(p_{t\mathbf{b}}, x_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$.[20] Then the market t-specific average utility of bundle $\mathbf{b}$ is $\delta_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} (\delta_j + x_{tj}\beta - \alpha p_{tj} + \xi_{tj}) + \Gamma_{t\mathbf{b}}$, with demand synergy:

$$\Gamma_{t\mathbf{b}} = \left(\delta_{\mathbf{b}} - \sum_{j \in \mathbf{b}} \delta_j \right) + \left(x_{t\mathbf{b}} - \sum_{j \in \mathbf{b}} x_{tj} \right) \beta - \alpha \left(p_{t\mathbf{b}} - \sum_{j \in \mathbf{b}} p_{tj} \right) + \left(\xi_{t\mathbf{b}} - \sum_{j \in \mathbf{b}} \xi_{tj} \right),$$

[19] Price endogeneity arises because the vector $(\xi_{tj})_{j=1}^{J}$ is observed by all price-setting firms but unobserved to the econometrician, while the dependence of $\ln(\mathscr{s}_{tj}/\mathscr{s}_{t0})$ on $\mathscr{s}_{t(j|j)}$ is typical of NL models (independently of price endogeneity), see Berry (1994).

[20] If $(p_{t\mathbf{b}}, x_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$ were unobserved, the endogenity problem discussed here would be even more severe, while the approach described in the next section would be unaffected.

where $\xi_{t\mathbf{b}}$ is an unobserved residual. Given these, system (9) can be re-written as:

$$\ln(\mathscr{s}_{t\mathbf{b}}) - \ln(\mathscr{s}_{t0}) = \delta_{\mathbf{b}} + x_{t\mathbf{b}}\beta - \alpha p_{t\mathbf{b}} + \ln\left(\sum_{j\in\mathbf{b}} \omega_{\mathbf{b}j}\left(\mathscr{s}_{t(\mathbf{b}|j)}\right)^{1-\lambda_j}\right) + \xi_{t\mathbf{b}}. \tag{13}$$

The term $\xi_{t\mathbf{b}}$ is a bundle-specific unobserved residual analogous to ξ_{tj} in (11). If one relied on (13) to simultaneously identify and estimate all the parameters of the PONL model, moment conditions (6) would not be sufficient for the bundle-level characteristics $(x_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$ to be exogeneous in each market t. In fact, such bundle-level exogeneity would require:

$$\begin{aligned} \mathbb{E}\left[(\xi_{tj})_{j=1}^{J}\,\middle|\,(x_{tj})_{j=1}^{J}, (x_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}\right] &= 0 \\ \mathbb{E}\left[(\xi_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}\,\middle|\,(x_{tj})_{j=1}^{J}, (x_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}\right] &= 0, \end{aligned} \tag{14}$$

where the first set of moment conditions in (14) already implies (6). Importantly, when moment conditions (14) do not hold, *all* the $K+J+1$ regressors in (13)—excluding the intercepts—will be endogenous, substantially complicating the task of finding a sufficient number of valid instruments. Differently, as we discuss next, none of the additional moment conditions in (14) is required for the exogeneity of $(x_{jt})_{j=1}^{J}$ in (11), so that the weaker moment conditions (6) plus the availability of $J+1$ valid instruments will suffice to addressing the endogeneity of p_{tj} and $\mathscr{s}_{t(j|j)}$ with respect to ξ_{jt}.

3.2 Identification from (11) and (12)

We now discuss the identification of $(\delta, \beta, \alpha, \lambda)$, with $\delta = (\delta_j)_{j=1}^{J}$ and $\lambda = (\lambda_j)_{j=1}^{J}$, and of $\Gamma_t = (\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$ in (11) and (12) from data on bundle-level purchase probabilities $(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$ and characteristics of single units $(x_{tj}, p_{tj})_{j=1}^{J}$ across T markets, with $T\to\infty$.[21] Note that, if the within-nest purchase probabilities $(\mathscr{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$ were observed, then identification would immediately follow from a sequential version of the classic instrumental variables argument by Berry (1994). One could first identify $(\delta, \beta, \alpha, \lambda)$ from linear regression (11) by instrumental variables (for the endogenous p_{tj} and $\mathscr{s}_{t(j|j)}$) and then Γ_t from nonlinear system (12) by a simple plug-in. However, the overlapping nesting structure of the PONL model prevents the observability of the within-nest purchase probabilities $(\mathscr{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$, which in turn leads to a different identification and estimation strategy.

Example 2. We illustrate the lack of observability due to the overlapping nests by slightly modifying Example 1 and adding bundle $(1,3)$ to the choice set $\mathbf{C}_1$. The NL model would require to uniquely allocate each element of $\mathbf{C}_1$ to a nest. Suppose that we specified three nests: $\mathbf{N}_i = \{1, (1,1), (1,2)\}$, $\mathbf{N}_{ii} = \{2, (2,2)\}$, and $\mathbf{N}_{iii} = \{3, (1,3)\}$. Then, given $(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$, one could directly obtain each within-nest purchase probability as $\mathscr{s}_{t(\mathbf{b}|g)} = \mathscr{s}_{t\mathbf{b}}/(\sum_{\mathbf{b}'\in\mathbf{N}_g}\mathscr{s}_{t\mathbf{b}'})$, $g = i, ii, iii$. Differently, because in the PONL

[21] Song and Chintagunta (2006), Sher and Kim (2014), Allen and Rehbeck (2019), and Wang (2019) study a different identification problem, where only the product-level purchase probabilities (marginals over bundles) are observed, rather than the bundle-level purchase probabilities.

model some $\mathbf{b}$ belongs to multiple nests, one cannot determine the within-nest purchase probabilities from the observed $(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$.[22] In the current example, we would have three overlapping nests: $\mathbf{N}_1 = \{1,(1,1),(1,2),(1,3)\}$, $\mathbf{N}_2 = \{2,(2,2),(1,2)\}$, and $\mathbf{N}_3 = \{3,(1,3)\}$. This leads to a system with 8 observed purchase probabilities and 9 unknowns:

$$
\begin{aligned}
\mathscr{s}_{tk} =& \mathscr{s}_{t(k|k)}\mathscr{s}_t^k \qquad k = 1,2,3 \\
\mathscr{s}_{t(j,j)} =& \mathscr{s}_{t(j,j|j)}\mathscr{s}_t^j \quad\ j = 1,2 \\
\mathscr{s}_{t(1,2)} =& \mathscr{s}_{t(1,2|1)}\mathscr{s}_t^1 + (1 - \mathscr{s}_{t(2|2)} - \mathscr{s}_{t(2,2|2)})\mathscr{s}_t^2 \\
\mathscr{s}_{t(1,3)} =& (1 - \mathscr{s}_{t(1|1)} - \mathscr{s}_{t(1,1|1)} - \mathscr{s}_{t(1,2|1)})\mathscr{s}_t^1 + (1 - \mathscr{s}_{t(3|3)})\mathscr{s}_t^3 \\
\mathscr{s}_{t0} =& 1 - \sum_{k=1}^{3} \mathscr{s}_t^k,
\end{aligned}
\tag{15}
$$

which would prevent the determination of the within-nest purchase probabilities. ■

In this context, identification can be achieved following Berry and Haile (2014) given the availability of valid instruments for p_{tj} and $\mathscr{s}_{t(j|j)}$. While this is standard, the associated derivations are useful to understand how to select valid instruments in practice (as discussed in the next section) and how to obtain a computationally convenient estimator (as discussed in section 4). Start by defining

$$
\pi_{tj} = \mathscr{s}_t^j/\mathscr{s}_{t0} = \left[\sum_{\mathbf{b}'\in\mathbf{N}_j} (\omega_{\mathbf{b}'j}\exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j}\right]^{\lambda_j}
$$

and by plugging $\delta_{tj} = \lambda_j \ln[\mathscr{s}_{tj}/\mathscr{s}_{t0}] + (1-\lambda_j)\ln\pi_{tj}$ in (12), so to obtain:

$$
\begin{aligned}
\Gamma_{t\mathbf{b}} &= \Gamma_{\mathbf{b}}(\Gamma_{t\mathbf{b}}; \pi_t, \lambda, \mathscr{s}_t) \\
&= \ln[\mathscr{s}_{t\mathbf{b}}/\mathscr{s}_{t0}] - \sum_{j\in\mathbf{b}} (\lambda_j \ln[\mathscr{s}_{tj}/\mathscr{s}_{t0}] + (1-\lambda_j)\ln\pi_{tj}) \\
&\quad - \ln\left(\sum_{j=1}^{J} \exp\left(\frac{\Gamma_{t\mathbf{b}}(1-\lambda_j)}{\lambda_j}\right)(\omega_{\mathbf{b}j})^{\frac{1}{\lambda_j}} \pi_{tj}^{1-\frac{1}{\lambda_j}} \prod_{r\in\mathbf{b}} [\mathscr{s}_{tr}/\mathscr{s}_{t0}]^{\frac{\lambda_r(1-\lambda_j)}{\lambda_j}} \pi_{tr}^{\frac{(1-\lambda_r)(1-\lambda_j)}{\lambda_j}}\right),
\end{aligned}
\tag{16}
$$

where $\pi_t = (\pi_{tj})_{j=1}^{J}$. Then, using $\delta_{t\mathbf{b}'} = \sum_{j\in\mathbf{b}'} \delta_{tj} + \Gamma_{t\mathbf{b}'}$, plug $\delta_{tj} = \lambda_j \ln[\mathscr{s}_{tj}/\mathscr{s}_{t0}] + (1-\lambda_j)\ln\pi_{tj}$ in the definition of π_{tj} and obtain:

$$
\begin{aligned}
\pi_{tj} &= \phi_j(\pi_t; \Gamma_t, \lambda, \mathscr{s}_t) \\
&= \left[\sum_{\mathbf{b}'\in\mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j} \exp(\Gamma_{t\mathbf{b}'}/\lambda_j) \prod_{r\in\mathbf{b}'} \left[\frac{\mathscr{s}_{tr}}{\mathscr{s}_{t0}}\right]^{\lambda_r/\lambda_j} \pi_{tr}^{(1-\lambda_r)/\lambda_j}\right]^{\lambda_j}.
\end{aligned}
\tag{17}
$$

Note that, given λ and $\mathscr{s}_t$, (16) and (17) define a system of C_1 equations and C_1 unknowns (Γ_t and

[22] Obviously, if there is no $\mathbf{b}$ that belongs to at least two nests (i.e., each $\mathbf{b} = (j, ..., j)$ only contains multiple units of a same product j), the PONL model simplifies to a standard NL whose nests partition $\mathbf{C}_1$.

π_t) for each t. Because each within-nest purchase probability $\mathscr{s}_{t(\mathbf{b}|j)}$ is a function of Γ_t and π_t (see the last equality in (7)), one can then address the lack of observability of $(\mathscr{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$ by expressing Γ_t and π_t in terms of λ and $\mathscr{s}_t$. To summarize, PONL model (11) and (12) implies:

$$
\begin{aligned}
\ln \mathscr{s}_{tj} - \ln \mathscr{s}_{t0} &= \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j)\ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}}\right) + \xi_{tj}, \text{ subject to} \\
\Gamma_t &= (\Gamma_{\mathbf{b}}(\Gamma_{t\mathbf{b}}; \lambda, \pi_t, \mathscr{s}_t))_{\mathbf{b}\in\mathbf{C}_2} \text{ from (16)} \\
\pi_t &= (\phi_j(\pi_t; \lambda, \Gamma_t, \mathscr{s}_t))_{j\in\mathbf{J}} \text{ from (17).}
\end{aligned}
\tag{18}
$$

While the presence of constraints (16) and (17) complicates estimation, it basically does not affect identification, in that (18) is subject to the same endogeneity concerns as (11): both p_{tj} and $\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}}$ are functions of the unobserved residuals $\xi_t = (\xi_{tj})_{j=1}^J$. Suppose that a vector of Q instruments z_{tj} with $Q \geq J+1$ is available, and that they satisfy the following moment conditions:

$$
\mathbb{E}\left[\xi_{tj}|z_{tj} = z\right] = 0, \text{ for all } j \in \mathbf{J} \text{ and } z \in \mathbf{D}_z, \tag{19}
$$

where $\mathbf{D}_z$ is the support of z. The next result confirms that the PONL model is identified on the basis of (18), the exogeneity of $(x_{tj})_{j=1}^J$, and the availability of instruments z_{tj} that satisfy (19).[23]

Proposition 1 (Identification). *Suppose that moment conditions* (6) *and Assumption 1 in Appendix A.1 hold. Then* $(\delta, \beta, \alpha, \lambda)$, Γ_t, *and* π_t *are identified for all* $t \in \mathbf{T}$.

Proof. See Appendix A.1. □

This shows that standard instrumental variables z_{tj} that satisfy (19) are sufficient not only to identify $(\delta, \beta, \alpha, \lambda)$, but also (Γ_t, π_t) for $t = 1, ..., T$ with $T \to \infty$. One can prevent any incidental parameter problem by relying on constraints (16) and (17) to concentrate out (Γ_t, π_t) given $(\lambda, \mathscr{s}_t)$ for each t. As a result, identification of the entire PONL model (including every (Γ_t, π_t)) boils down to the unique determination of $(\delta, \beta, \alpha, \lambda)$ by instrumental variables from a nonlinear system. This is important because the estimation of price elasticities and marginal costs, and the simulation of counterfactuals (e.g., alternative pricing strategies and mergers) usually require knowledge of the entire model.

3.3 Choice of Instruments

Proposition 1 demonstrates that the PONL model is identified on the basis of (19) given the availability of at least $1 + J$ valid instruments. While theoretically reassuring, this result is silent about how to practically choose them among the several possible alternatives appeared in the literature (Berry and Haile, 2016; Gandhi and Houde, 2019). In this section, we complement Proposition 1 and provide some practical guidance on the selection of valid instruments.

[23] Formally, following Berry and Haile (2014), Proposition 1 relies on the completeness condition embedded in Assumption 1 (as detailed in Appendix A.1) rather than on condition (19). More practically, however, its essence can be summarized by the availability of at least $J + 1$ valid instruments, as in (19).

Different categories of instruments were proposed in the literature to address, respectively, the endogeneity of p_{tj} and that of $\frac{s_{tj}/s_{t0}}{\pi_{tj}}$ in (18) (Berry and Haile, 2016; Gandhi and Houde, 2019). Classical instruments for p_{tj} are excluded cost-shifters (e.g., input prices) or, when these are not available, some proxies for these or for marginal costs. For example, Hausman (1996) proposed to proxy the marginal cost of a product in a specific market by the price of the same product from different markets, relying on the idea that the marginal cost of a product should be similar across markets (Nevo, 2001). As popularized by Berry et al. (1995), other classic instruments for p_{tj} are the exogenous characteristics x_{tk} for any product $k \neq j$, with the idea that more or less substitutability in characteristic space should lead to more or less price competition among products.

Despite the lack of observability of π_{tj}, appropriate instruments for $\frac{s_{tj}/s_{t0}}{\pi_{tj}}$ can be selected on the basis of their correlation with $s_{t(j|j)}$. To see this, denote by $\pi_t = (\pi_j(\lambda; s_t))_{j \in \mathbf{J}}$ a solution to constraints (16) and (17) for given λ and s_t.[24] By a first-order Taylor approximation of $\ln(\pi_{tj}) = \ln(\pi_j(\lambda; s_t))$ in (18) around its true value $\ln(\pi^0_{tj}) = \ln(\pi_j(\lambda^0; s_t))$, we obtain:

$$\begin{aligned}\ln s_{tj} - \ln s_{t0} &= \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j)\left[\ln\left(\frac{s_{tj}/s_{t0}}{\pi^0_{tj}}\right) - \frac{1}{s^j_t}\frac{\partial \pi_j(\lambda^0; s_t)}{\partial \lambda}(\lambda - \lambda^0)\right] + \xi_{tj} \\ &= \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j)\left[\ln s_{t(j|j)} - \frac{1}{s^j_t}\frac{\partial \pi_j(\lambda^0; s_t)}{\partial \lambda}(\lambda - \lambda^0)\right] + \xi_{tj},\end{aligned} \tag{20}$$

where the leading term of the first-order Taylor expansion is $\ln s_{t(j|j)}$.[25] As a consequence, and similar to a scenario in which π_{tj} were observed, a valid instrument here is "something" that shifts $s_{t(j|j)}$ independently of ξ_{tj}, therefore helping to identify $1 - \lambda_j$ (Berry, 1994). From (7), we can also re-express $s_{t(j|j)}$ as:

$$s_{t(j|j)} = \frac{\exp(\delta_{tj})^{1/\lambda_j}}{\sum_{\mathbf{b}' \in \mathbf{N}_j}(\omega_{\mathbf{b}'j}\exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j}} = \frac{1}{1 + \sum_{\mathbf{b}' \in \mathbf{N}_j, \mathbf{b}' \neq j}(\omega_{\mathbf{b}'j}\exp(\delta_{t\mathbf{b}'} - \delta_{tj}))^{1/\lambda_j}}. \tag{21}$$

Given the last two equations, $\delta_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} \delta_{tj} + \Gamma_{t\mathbf{b}}$, moment conditions (6), and the overlapping nesting structure of the PONL model, it is simple to construct appropriate instruments for $\frac{s_{tj}/s_{t0}}{\pi_{tj}}$ relying both on product-level and bundle-level exogenous characteristics. For instance, the characteristics of product k, x_{tk} with $k \neq j$, will be valid product-level instruments for $\frac{s_{tj}/s_{t0}}{\pi_{tj}}$ as long as the nests j and k are overlapping, $\mathbf{N}_j \cap \mathbf{N}_k \neq \emptyset$ (there exist bundles including at least a unit of both j and k). Moreover, if one observes bundle-level characteristics $x_{t\mathbf{b}} \neq \sum_{k \in \mathbf{b}} x_{tk}$ and is willing to additionally assume the first set of moment conditions in (14), then (21) implies that $x_{t\mathbf{b}'} - x_{tj}$ is a valid bundle-level instrument given its correlation with $s_{t(j|j)}$ through $\delta_{t\mathbf{b}'} - \delta_{tj}$ (Gandhi and Houde, 2019). As a special case, note that in applications with $x_{t\mathbf{b}} = \sum_{k \in \mathbf{b}} x_{tk}$, such as the one we study in this paper,

[24] By substituting (16) into (17), for given λ and s_t, π_t is the only argument of the resulting system. As a consequence, we can express π_t in terms of λ and s_t. See Assumption 2 in Appendix B for more details.

[25] Note that the second term inside the square brackets is proportional to $\lambda - \lambda^0$. Its contribution to pin down λ_j will then diminish when λ is close enough to λ^0, i.e. when the estimator of λ^0 is consistent. Consequently, in practice, the main channel to identify λ_j will be the leading term $\ln s_{t(j|j)}$.

moment conditions (6) are sufficient also for the validity of this type of bundle-level instruments.

Through a similar mechanism, also instruments for *excluded* prices (i.e., all but p_{tj}) can be valid for $\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_{tj}}$: for example, any excluded cost-shifter for product $k \neq j$ such that $\mathbf{N}_j \cap \mathbf{N}_k \neq \emptyset$ would affect $\mathcal{s}_{t(j|j)}$ through p_{tk} independently of ξ_{tj}. The validity of the instruments for price to addressing also endogeneity of the within-nest purchase probability is specific to the exclusion restrictions embedded in (20) and (21). In more general demand models than PONL, endogeneity of the purchase probabilities calls for a source of exogenous variation independent of prices (Berry and Haile, 2016). In such general models, when additional data are not readily available, the exogenous characteristics may be the only valid instruments to addressing endogeneity of the purchase probabilities (Berry and Haile, 2014).

4 Estimation

Given data on bundle-level purchase probabilities $(\mathcal{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$, a natural approach to estimating $(\delta, \beta, \alpha, \lambda)$ and $(\Gamma_{t\mathbf{b}})_{t\in\mathbf{T},\mathbf{b}\in\mathbf{C}_2}$ is the Generalized Method of Moments (GMM) estimator proposed by Berry et al. (1995).[26] This could be obtained on the basis of purchase probabilities (7)-(8) and moment conditions (14), and then relying either on the fixed point approach (Aguirregabiria and Mira, 2002; Berry et al., 1995; Rust, 1987) or on the MPEC approach (Dubé et al., 2012; Su and Judd, 2012) for implementation. Unfortunately, this GMM estimator would be impractical with large choice sets, mainly because of the large dimensionality $T \times C_1$ of the demand system. In particular, the fixed point implementation of Berry et al. (1995) would require the computation of T separate demand inverses with C_1 equations each (and no readily available contraction mapping results), while the MPEC implementation would require the computation of up to $T \times C_1$ nonlinear constraints and their derivatives.[27]

To overcome this challenge, we mimic our identification strategy and propose a Constrained Two Stage Least Square (C2SLS) estimator on the basis of (18). The proposed C2SLS estimator is a natural extension of the Two Stage Least Square (2SLS) estimator by Berry (1994) to the case of unobserved within-nest purchase probabilities (arising from the overlapping nests). Importantly, the C2SLS can be implemented by a convenient iterative procedure that is optimization- and derivative-free, and parallelizable over both bundles and markets, virtually eliminating any challenge of dimensionality due to large choice sets. We show that the C2SLS estimator has desirable asymptotic properties and that, upon numerical convergence, the proposed iterative procedure always implements it.

[26] Given the lack of observability of the within-nest purchase probabilities $(\mathcal{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$, one cannot directly construct an estimator on the basis of (9) as for NL models with non-overlapping nests (Berry, 1994), but must rely on the more general approach by Berry et al. (1995).

[27] See footnote 4.

4.1 A Constrained Two Stage Least Square (C2SLS) Estimator

Our constructive identification strategy readily leads to a C2SLS estimator based on (18) and instruments (19), as a solution to the following nonlinear system:

$$\begin{cases} (\delta, \beta, \alpha, 1-\lambda) = \left(X^{\mathrm{T}}(ZZ^{\mathrm{T}})X\right)^{-1}\left(X^{\mathrm{T}}(ZZ^{\mathrm{T}})Y\right) \\ \text{Nonlinear equations (16): } \Gamma_t = \left(\Gamma_{\mathbf{b}}(\Gamma_{t\mathbf{b}}; \lambda, \pi_t, \mathscr{s}_t)\right)_{\mathbf{b}\in\mathbf{C_2}} \\ \text{Nonlinear equations (17): } \pi_t = \left(\phi_j(\pi_t; \lambda, \Gamma_t, \mathscr{s}_t)\right)_{j\in\mathbf{J}} \end{cases} \tag{22}$$

where $Z = (Z_t)_{t=1}^T$, $X = \left((e_j)_{j=1}^J, x_t, -p_t, \left(\ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}}\right)\right)_{j=1}^J\right)_{t=1}^T$, e_j is a vector of zeros with j^{th} element equal to 1, and $Y = \left((\ln \mathscr{s}_{tj} - \ln \mathscr{s}_{t0})_{j=1}^J\right)_{t=1}^T$. Denote by $(\delta^0, \beta^0, \alpha^0, \lambda^0)$ and $(\pi_t^0, \Gamma_t^0)_{t=1}^T$ the true parameter values. The C2SLS in (22) is a natural extension of the 2SLS proposed by Berry (1994) for NL models with non-overlapping nests. To see this, suppose that the within-nest purchase probabilities $(\mathscr{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$ were observed, as in the classic NL. Then one could, first, estimate $(\delta, \beta, \alpha, \lambda)$ by 2SLS as a solution to the *linear* equations in (22) and, second, estimate Γ_t from nonlinear system (12) by a plug-in. Differently, with overlapping nests one does not observe the within-nest purchase probabilities, but however knows that they must satisfy both the *linear* (as the 2SLS) and the *nonlinear* equations in (22), giving rise to the C2LS estimator.

Proposition 2 (Asymptotic Properties)**.** *Suppose Assumptions 2 and 3 in Appendix B hold.*

- *A solution to* (22) *in a neighbourhood of* $(\delta^0, \beta^0, \alpha^0, \lambda^0)$ *and* $(\pi_t^0, \Gamma_t^0)_{t=1}^T$*, denoted by* $(\hat\delta, \hat\beta, \hat\alpha, \hat\lambda)$ *and* $(\hat\pi_t, \hat\Gamma_t)_{t=1}^T$*, exists with probability one as* $T \to \infty$*.*

- $(\hat\delta, \hat\beta, \hat\alpha, \hat\lambda)$ *and* $(\hat\pi_t, \hat\Gamma_t)_{t=1}^T$ *is consistent and asymptotically normal.*

Proof. See Appendix B. □

The first result of Proposition 2 confirms that the C2SLS estimator is well defined and always exists in large samples, while the second guarantees that it has desirable asymptotic properties. In Appendix B.1, we derive the asymptotic variance-covariance matrix and a simple plug-in procedure to compute it. Even though well behaved in theory, the C2SLS estimator can be challenging to implement, especially with large choice sets. While the 2SLS by Berry (1994) only solves the linear equations in (22), the C2SLS requires the solution of the entire nonlinear system (22). With large C, the nonlinear part of system (22) introduces practical complexities not present in the 2SLS: in addition to $(\delta, \beta, \alpha, \lambda)$, one also needs to compute $T \times J$ values of $(\pi_t)_{t=1}^T$ and $T \times C_2$ values of $(\Gamma_t)_{t=1}^T$ that simultaneously satisfy (16) and (17). We circumvent this computational challenge by proposing an iterative procedure that does not attempt the direct numerical solution of nonlinear system (22), but only executes a sequence of 2SLS estimators and parallelizable plug-in operations. Together, these simple steps largely reduce the computational time and memory requirements needed to implement the C2SLS estimator.

4.2 A Convenient Iterative Procedure

Denote the algorithm's iterations by $k = 1, ..., \bar{K}$ and the parameter values obtained at iteration k by superscript (k). Given starting values $\left(\delta^{(0)}, \beta^{(0)}, \alpha^{(0)}, \lambda^{(0)}\right)$ and $\left(\pi_t^{(0)}, \Gamma_t^{(0)}\right)_{t \in \mathbf{T}}$, at each iteration k execute the following steps:[28]

Step 1. Given $\pi_t^{(k-1)}$, $\lambda^{(k-1)}$, and $\Gamma_t^{(k-1)}$, for each (t, j) compute $\pi_{tj}^{(k)}$ as a plug-in from the right-hand side of (17).

Step 2. Given $\pi_t^{(k)}$, compute $\left(\delta^{(k)}, \beta^{(k)}, \alpha^{(k)}, \lambda^{(k)}\right)$ by 2SLS from the linear equations in (22), i.e. ignoring nonlinear equations (16) and (17).

Step 3. Given $\pi_t^{(k)}$, $\lambda^{(k)}$, and $\Gamma_t^{(k-1)}$, for each $(t, \mathbf{b})$—independently of any other market and bundle—compute $\Gamma_{t\mathbf{b}}^{(k)}$ as a one-step Newton-Raphson approximation to the unique solution of (16).[29]

Step 4. If $k < \bar{K}$, move on to the next iteration $k + 1$. If instead $k = \bar{K}$, exit the algorithm.

Step 2 of the proposed algorithm leverages on the observation that, for any given value of π_t, the linear equations in (22) are nothing but the 2SLS popularized by Berry (1994) for the estimation of NL models. Then, steps 1 and 3 simply *update* the values of $(\Gamma_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$ and π_t rather than *fully solving* nonlinear equations (16) and (17). From a computational perspective, each of these steps is very convenient. Step 2 only requires the estimation of a 2SLS, while steps 1 and 3 consist of a fully parallelizable sequence of plug-ins (which do not involve any numerical optimization or derivative).[30]

The proposed algorithm mimics the classic Gauss-Seidel method for the solution of linear systems to implementing the C2SLS estimator, a solution of nonlinear system (22). While similar algorithms were shown to practically facilitate the implementation of linear (Guimaraes and Portugal, 2010) and nonlinear fixed effects estimators (Hospido, 2012), little is known about their numerical convergence: despite the practical convenience, it is often unclear whether these algorithms numerically converge to the desired estimators. Importantly, the next result establishes that whenever our simple sequence of regressions and plug-ins numerically converges, it will attain the C2SLS estimator.

Proposition 3 (Numerical Convergence)**.** *Suppose that for all $t = 1, ..., T$ and $\mathbf{b} \in \mathbf{C}_2$, as $\bar{K} \to \infty$, $\pi_t^{(\bar{K})} \to \pi_t^*$ and $\Gamma_{t\mathbf{b}}^{(\bar{K})} \to \Gamma_{t\mathbf{b}}^*$ for some $\pi_t^* \in \mathbb{R}^J$ and $\Gamma_{t\mathbf{b}}^* \in \mathbb{R}$. Then, as $\bar{K} \to \infty$, $\left(\delta^{(\bar{K})}, \beta^{(\bar{K})}, \alpha^{(\bar{K})}, \lambda^{(\bar{K})}\right)$ and $\left(\pi_t^{(\bar{K})}, \Gamma_t^{(\bar{K})}\right)_{t=1}^{T}$ converge to the C2SLS estimator.*

Proof. See Appendix D. □

[28] While here we only sketch the main features of the proposed iterative procedure, in Appendix C we discuss several implementation details: from the choice of starting values (parameter values at iteration 0) and stopping criteria ($\bar{K}$), to the updating in steps 1 and 3.

[29] Importantly, as shown in Lemma 1, Appendix A.1, (16) has a unique solution $\Gamma_{t\mathbf{b}}$ which is independent of any other market and bundle other than $(t, \mathbf{b})$.

[30] As discussed in Appendix C, the updating in step 3, despite being a one-step Newton-Raphson approximation, does not require any numerical differentiation: the derivatives of (16) have a simple analytical form.

This guarantees that the convergence of each $\pi_t^{(\bar{K})}$ and $\Gamma_{t\mathbf{b}}^{(\bar{K})}$ can only happen to the C2SLS estimator. To test if the algorithm has implemented the C2SLS estimator, it suffices to verify whether the iterative procedure has numerically converged.[31] Even though Proposition 3 does not guarantee the numerical convergence of the proposed algorithm, and thus its ability to produce the C2SLS estimates, reassuringly, in the large number of estimates we performed between the Monte Carlo simulations and the empirical application, we never experienced any lack of numerical convergence. In the hypothetical case of lack of numerical convergence, we suggest to re-launch the algorithm from different starting values (as typically done for validation in analogous numerical procedures, Robert and Casella, 2013).

4.3 Discussion

In the context of the PONL model, the C2SLS estimator in (22) is simpler to implement than the classic GMM estimator by Berry et al. (1995). The greater simplicity comes from the possibility to fully parallelize the estimation of the $T \times C_2$ demand synergies, which can be performed for each $\Gamma_{t\mathbf{b}}$ in isolation. This independence would not be exploited in the classic GMM estimator by Berry et al. (1995), which would instead compute an inverse of the entire demand system (C_1 equations) within each market (so parallelizing only over markets but not over bundles). Despite this computational advantage, as discussed above, a standard implementation of the C2SLS estimator by directly solving (22) could still be problematic with large C. However, the proposed iterative procedure virtually eliminates any challenge of dimensionality related to large choice sets.

An essential factor behind the practical advantages of the proposed C2SLS estimator is the use of individual-level purchases in the aggregate form of bundle-level purchase probabilities. Bundle-level purchase probabilities are not typically directly observed (with the exception of a few industries, see Crawford and Yurukoglu, 2012; Song et al., 2017) but rather computed from samples of individual-level purchases (Ershov et al., 2021) and thus subject to sampling error. When the number of bundles is large relative to the sample of individual-level purchases, sampling error in the bundle-level purchase probabilities can be pronounced and lead to estimation bias (Gentzkow et al., 2019), for example because of the large number of observed "zeros" (Gandhi et al., 2020). Even though, in the interest of space, we do not address this complication in the current paper, the C2SLS estimator can be extended to control for sampling error in the bundle-level purchase probabilities by building on the de-biasing technique proposed by Freyberger (2015).

Following a different route, one could instead opt for more traditional likelihood-type estimators based on the direct use of individual-level purchases (Aryal and Gabrielli, 2020; Gentzkow, 2007; Grzybowski and Verboven, 2016; Iaria and Wang, 2019; Kwak et al., 2015; Ruiz et al., 2020). Unfortunately,

[31] In practice, numerical convergence is usually defined by a stopping criterion, such as that the distance in the parameter values between two consecutive iterations is smaller than a threshold. For instance, in our simulations and empirical application, we consider the algorithm to have converged when the absolute values of $\Gamma_{t\mathbf{b}}^{(k)} - \Gamma_{t\mathbf{b}}^{(k-1)}$, $\pi_{tj}^{(k)} - \pi_{tj}^{(k-1)}$, and $\left(\delta^{(k)}, \beta^{(k)}, \alpha^{(k)}, \lambda^{(k)}\right) - \left(\delta^{(k-1)}, \beta^{(k-1)}, \alpha^{(k-1)}, \lambda^{(k-1)}\right)$ are small enough for all t and $\mathbf{b}$. As shown in section 5, our Monte Carlo simulations suggest that 5 iterations can already be sufficient to achieve this form of numerical convergence.

also this approach would not be computationally convenient with large choice sets, mainly because of the large number of fixed effects required to control for price endogeneity (Iaria and Wang, 2019). In this sense, in applications with large choice sets, the proposed iterative procedure may be the only practically viable estimation alternative.

5 Monte Carlo Simulations

We now present simulation results to investigate the finite sample performance of the C2SLS estimator as a function of both the choice set size C and the number of iterations $\bar{K}$ in the proposed algorithm.

5.1 Data Generating Process

We generate data from a PONL model with $J = 10$ products and bundles that combine multiple units these. Across experiments, we vary the maximum "dimension" of the bundles included in the choice set: the maximum number of units that can be jointly purchased as a bundle, and consequently the size of the choice set. For example, with bundles of dimension up to two, individuals can choose among 66 bundles of the kind (j_1, j_2), with $j_k \in \{0\} \cup \mathbf{J}$, $k = 1, 2$; while with bundles of dimension up to three, individuals can instead choose among 286 bundles of the type (j_1, j_2, j_3) with $j_k \in \{0\} \cup \mathbf{J}$, $k = 1, 2, 3$.[32] In particular, we consider choice sets of increasing size, with $C \in \{66, 286, 1001, 3003, 8008, 19448\}$, where 66 is the number of bundles of dimension up to two, 286 the number of bundles of dimension up to three, and so on until 19448, the number of bundles of dimension up to seven.

We specify $((\delta_j = 1, \lambda_j = 0.4)_{j=1}^{10}, \beta = 2, \alpha = 2)$, the demand synergies as $\Gamma_{t\mathbf{b}} \sim N(0, 0.1)$, the product-specific unobserved residuals as $\xi_{tj} \sim N(0, 0.2)$, the product-specific exogenous characteristic as $\log(x_{tj}) \sim N(1, 0.1)$ (we set $K = 1$ for simplicity), and the product-specific marginal cost as $\log(z_{tj}) \sim N(1, 0.1)$. We assume that in each market t, a monopolist sets the unit prices of the 10 products (independently across markets), $(p_{tj})_{j=1}^{10}$, and linear pricing: the price of each bundle $\mathbf{b}$ is given by the sum of the unit prices of the units it includes, $p_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} p_{tj}$. We assume that the monopolist faces no technological advantage or disadvantage in selling bundles: the marginal cost of any bundle $\mathbf{b}$ is given by the sum of the marginal costs of the units it includes, $z_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} z_{tj}$.

Despite assuming $\delta_j = \delta_k$ and $\lambda_j = \lambda_k$ for any $j \neq k$ in the data generating process, we do not impose such constraint in estimation and allow for different δ_j and λ_j for each $j = 1, ..., 10$. We use two types of instruments to deal with the endogeneity of p_{tj} and $\frac{s_{tj}/s_{t0}}{\pi_{tj}}$. First, we instrument price p_{tj} by polynomials of the exogenous characteristic x_{tj} and of the marginal cost z_{tj}. Second, we instrument $\frac{s_{tj}/s_{t0}}{\pi_{tj}}$ by weighted averages of $x_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} x_{tj}$ and of $z_{t\mathbf{b}}$ among the bundles that belong to $\mathbf{N}_j \setminus \{j\}$.[33]

[32] For example, the 66 bundles in the case of bundles of dimension up to two are: the choice of not purchasing any product $(0, 0)$, the 10 single units of the products in $\mathbf{J}$, and the 55 bundles of dimension two in $\mathbf{C}_2 = \mathbf{J} \times \mathbf{J}$. Note that we allow for the purchase of bundles that include multiple units of the same product.

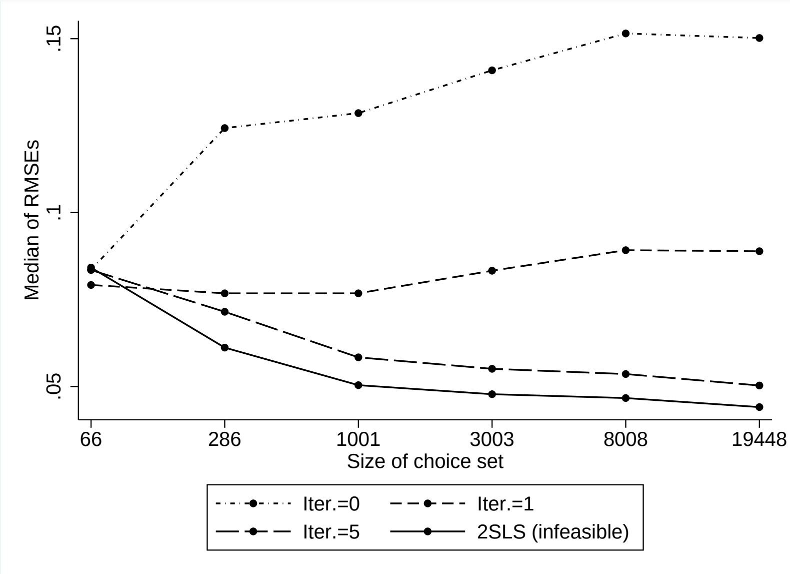
[33] More precisely, we use the following list of instruments: $Z_{tj} = \Big(e_j^{\mathrm{T}}, x_{tj}, x_{tj}^2, x_{tj}^3, z_{tj}, z_{tj}^2, z_{tj}^3, x_{tj}z_{tj}, x_{tj}^2 z_{tj}, x_{tj}z_{tj}^2, \left(\sum_{\mathbf{b} \in \mathbf{N}_j \setminus \{j\}} \omega_{\mathbf{b}j} x_{t\mathbf{b}}\right) e_j^{\mathrm{T}}, \left(\sum_{\mathbf{b} \in \mathbf{N}_j \setminus \{j\}} \omega_{\mathbf{b}j} z_{t\mathbf{b}}\right) e_j^{\mathrm{T}}\Big)^{\mathrm{T}}$, where e_j is a vector of zeros with j^{th} element equal to 1.

5.2 Simulation Results

We compare the finite sample performance of the proposed iterative procedure for the C2SLS estimator in (22) with respect to the infeasible two-step procedure that relies on the observability of the within-nest purchase probabilities, i.e. first estimating the $2J + K + 1 = 22$ parameters in (11) by 2SLS and then each of the $C_2 \times T$ demand synergies by an independent plug-in as in (12). In terms of performance, the infeasible two-step procedure is an upper bound for the C2SLS, which estimates the same parameters but without relying on the observability of the within-nest purchase probabilities.

Figure 2: Median of RMSEs of C2SLS Estimator

(a) Parameters $(\delta, \beta, \alpha, \lambda)$, $T = 200$ markets



(b) Choice Set Size $C = 286$

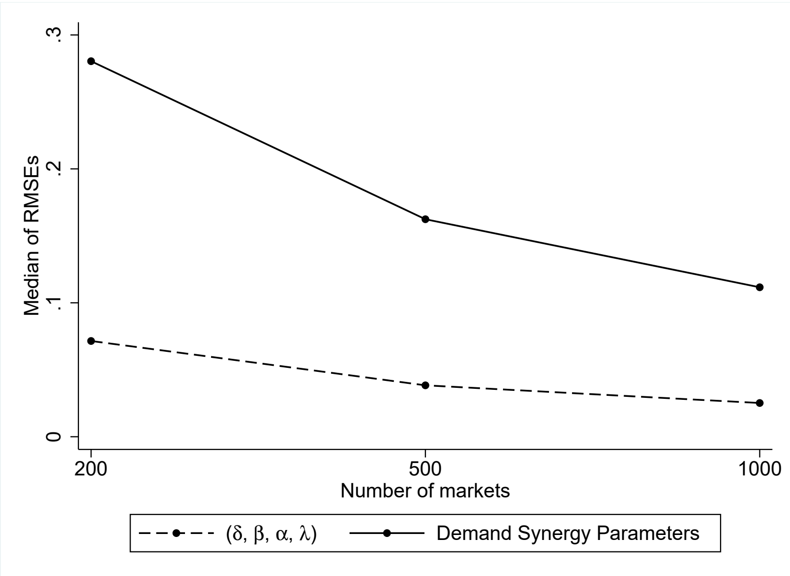


Figure 2 summarizes our simulation results. Figure 2(a) illustrates results for the estimation of $(\delta, \beta, \alpha, \lambda)$ in different scenarios with choice set size $C \in \{66, 286, 1001, 3003, 8008, 19448\}$. For each C, we simulate 100 datasets/repetitions with $T = 200$ markets and then average estimation results across these. We summarize the finite sample performance of each estimator in terms of its median Root Mean Square Error (RMSE).[34] The solid line represents the median RMSE of the infeasible 2SLS, while the others plot the median RMSE of the proposed algorithm after iteration 0 (see Appendix C for a detailed description of this starting iteration), iteration 1, and iteration 5.

Figure 2(a) shows how, in practice, the proposed iterative procedure converges very fast to the infeasible 2SLS estimator, the theoretical upper bound for the C2SLS estimator. After only five iterations, the median RMSE of the proposed algorithm is almost indistinguishable from that of the infeasible 2SLS estimator. Importantly, the fast convergence holds irrespectively of the choice set size C, confirming that a few iterations may be sufficient to implement the C2SLS estimator (Proposition 3) also in empirical applications with large choice sets.

Figure 2(b) illustrates results for the estimation of all parameters in various scenarios with a constant choice set size $C = 286$ (i.e., all bundles of size 3) but an increasing number of markets $T \in \{200, 500, 1000\}$. For each T, we simulate 100 datasets/repetitions and plot the median RMSE of the proposed algorithm after five iterations. The dashed line plots the median RMSE of $(\delta, \beta, \alpha, \lambda)$, while the solid line represents the median RMSE of the demand synergy parameters. While for any T the demand synergy parameters are less precisely estimated than $(\delta, \beta, \alpha, \lambda)$, a larger T—in line with Proposition 2—corresponds to a better performance of the C2SLS estimator.

6 Quantity Discounts and Carbonated Soft Drinks

We implement our methods to empirically investigate the determinants and welfare consequences of quantity discounts in the market for carbonated soft drinks in the USA. Relying on household-level purchase data from the period 2008-2011, we estimate a flexible PONL model and then assess the welfare effects of the observed quantity discounts by simulating a counterfactual with linear pricing.

6.1 Data, Definitions, and Descriptive Statistics

We use household-level and store-level IRI data on carbonated soft drinks (CSDs) for the cities of Pittsfield and Eau Claire (USA) in the period 2008-2011. We report a brief description of the data here and refer the reader to Bronnenberg et al. (2008) for more detail.

We focus on the $I = 6,155$ households who are observed to purchase CSDs at least once from 2008 until 2011. For these, we observe household size (number of family members) and a panel of $1,736,012$ household-level shopping trips to 22 different grocery stores over a period of 208 weeks. A

[34] For given C and estimator of θ, we compute the parameter-specific RMSE for each parameter $d = 1, ..., D$ in $\hat{\theta}_r = (\hat{\theta}_{1r}, ..., \hat{\theta}_{Dr})$ across $r = 1, ..., 100$ repetitions: $\text{RMSE}(\hat{\theta}_d, \theta_d) = \sqrt{\frac{1}{100}\sum_{r=1}^{100}(\hat{\theta}_{dr} - \theta_d)^2}$. We then plot the median of the parameter-specific $\text{RMSE}(\hat{\theta}_d, \theta_d)$ across the D parameters in θ.

Table 1: Descriptive Statistics

Product Definition	
Brand	variable "L5" in IRI
Producer	Coca-Cola, PepsiCo, Dr. Pepper, Others
Sample Characteristics	
Num. of UPCs	$1,683$
Num. of products	128
Num. of weeks	208
Num. of households	$6,155$
% single-person households	24.55%
Num. of shopping trips	$1,736,012$
% shopping trips with purchase	23.71%
Shopping frequency, any purchase	1.36 times per week
Shopping freq., with CSDs purchase	every 2.22 weeks
Num. of markets (four weeks $\times$ store)	$1,197$
Average num. shop. trips per market	$1,450.30$
Purchased Quantities (in units)	
Average per household-year	117.24
Single-person households, average per year	66.02
Multi-person households, average per year	133.91
Average bundle size (units per shopping trip)	6.99
Num. of bundles	$16,873$
Average num. of bundles per market	123.80
% shop. trips with multi. units (A+B)	93.24%
(A) multi. units same prod.	90.15%
(B) multi. units diff. prod.	9.85%

shopping trip is defined as a household's purchase occasion to a grocery store in a given day. Each shopping trip records all the Universal Product Codes (UPCs) purchased by a household across all product categories sold by the store: during 23.71% of these, CSDs are observed to be purchased. We consider a household to choose the outside option whenever no CSD product is purchased during a shopping trip (in general, something must be purchased for a shopping trip to be in the data).

We observe households to purchase CSDs on average every 2.22 weeks, which suggests that, on average, they deplete their stocks of accumulated CSDs in approximately two weeks. We then define a market $t = 1, ..., 1197$ as a (four weeks $\times$ store) combination to make sure that observed purchases correspond to consumption within the same interval of time. This mitigates concerns about stockpiling behavior in which households buy more "now" for "later" (Hendel and Nevo, 2006, 2013; Wang, 2015).

Households are observed to purchase $1,683$ different UPCs of CSDs mainly by three large producers, Coca-Cola, PepsiCo, and Dr. Pepper, plus some smaller ones we collectively label "Others." From these UPCs, we define products on the basis of the "brand" variable L5 in the IRI data (e.g., Coke Classic or Diet Pepsi), considering all the UPCs by Others as a single product. This results in 128 products, listed in Table 10, Appendix F.6. The top two panels of Table 1 summarize this information.

We discretize purchased quantities in units of one liter (L): we consider purchases of a product up to 1L as one unit, between 1L and 2L as two units, and so on until 154 units, the largest purchased quantity of a single CSD product during a shopping trip we observe in the data. We denote a bundle **b** as any combination of units of the same and of different CSDs we observe to be purchased during any one shopping trip.[35] For brevity, we refer to "units" or "liters" interchangeably and call "bundles" also the purchases of single units (i.e., bundles of size one). On average, we observe households to purchase 117.24 units of CSDs per year.[36]

As summarized in the bottom panel of Table 1, we observe $16,873$ different bundles to be purchased during any shopping trip in any market, with an average of 123.80 different bundles within each market. As is well known (Chan, 2006; Dubé, 2004; Ershov et al., 2021), the purchase of multiple units (6.99 on average) is a common phenomenon in the market for CSDs, which we observe in 93.24% of the shopping trips with any purchase of CSDs. In 9.85% of these, households purchase multiple units of different CSD products, stressing the importance of allowing for mixed bundles in the demand model. We divide households into two groups on the basis of their family size: single-person or multi-person, $hs \in \{\text{single, multi}\}$. Figure 3 shows that, as expected, multi-person households tend to purchase bundles of larger sizes than single-person households, both in terms of the same product (panel b) and of different products (panel c). Because of this, we allow for the possibility that households of different sizes react differently to quantity discounts: we compute choice probabilities conditional on hs and allow for different household sizes to have different parameters. In particular, we compute each $\jmath^{hs}_{t\mathbf{b}}$ as the proportion of shopping trips in t corresponding to purchases of **b** by households of size hs.

We compute each bundle-level price $p_{t\mathbf{b}}$ as the average observed price (in US dollars) across all shopping trips in t corresponding to purchases of **b**. Note that third-degree price discrimination cannot be implemented in this context and, within each market, households of different sizes face the same prices. For each UPC purchased during any shopping trip in a (week$\times$store) combination, IRI reports the average price in that (week$\times$store) from a separate store-level sales dataset. As a consequence, $p_{t\mathbf{b}}$ averages over (i) the purchased combinations of UPCs that correspond to the same **b** (given our definitions of product and quantity) and (ii) four consecutive weeks (given our definition of market). Because IRI records the average (week$\times$store) price of each UPC separately, we do not observe nonlinear prices across UPCs of different products (e.g., joint purchase of 2L Coke Classic and 2L Sprite) and instead focus on quantity discounts across UPCs involving different volumes of the same product (e.g., 1L Coke Classic versus 2L Coke Classic).

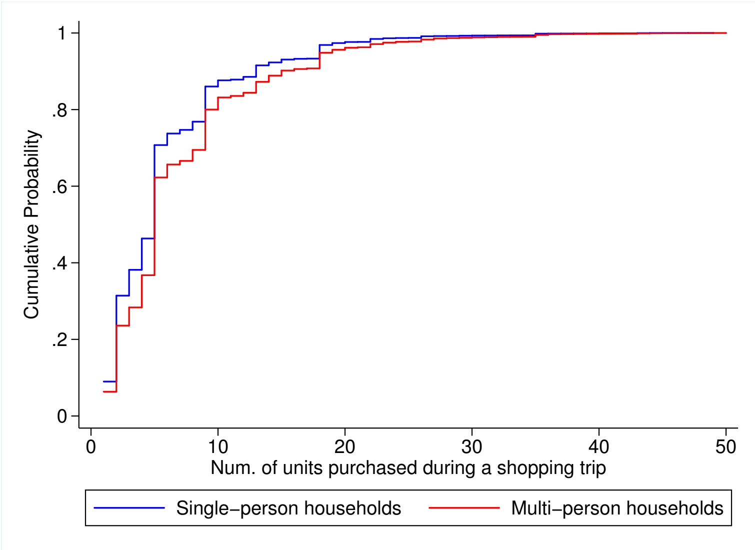
Table 2 provides descriptive evidence about the presence of quantity discounts. We regress the

[35]In some markets, some of the CSD products are only observed to be purchased through bundles and never in isolation as single units. Without further assumptions, the C2SLS estimator cannot pin down the demand synergy $\Gamma_{t\mathbf{b}}$ of bundles that include these products in such markets. In Appendix E, we discuss a simple procedure to extend the use of the C2SLS estimator to cases like this where some product can only be purchased through bundles.

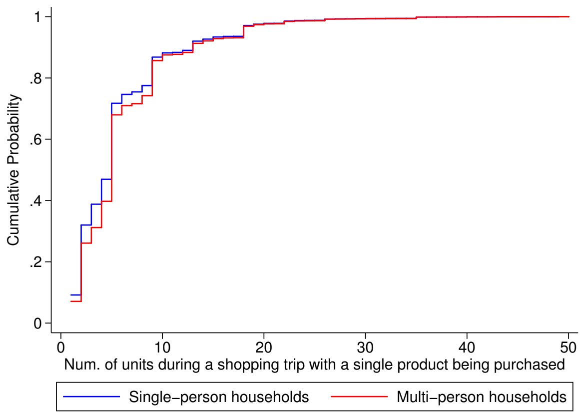
[36]The observed average purchased quantity of 117.24 units per year is smaller than the 156L reported by Allcott et al. (2019) on the basis of the Nielsen data (for the period 2007-2016). There are at least two possible explanations. First, the Nielsen household-level scanner data may cover a larger number of retailers than IRI, so that a larger share of purchases of CSDs is not recorded in our data. Second, the composition of demographics sampled by Nielsen and IRI may differ, so that Nielsen's households purchase larger quantities of CSDs.

Figure 3: Number of Purchased Units, Single- and Multi-Person Households

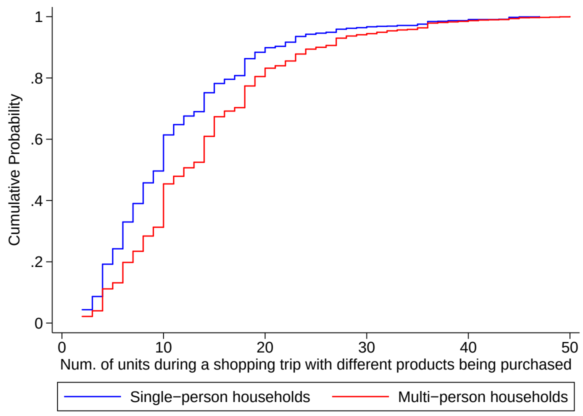
(a) All shopping trips



(b) Shopping trips with multiple units of same product



(c) Shopping trips with multiple units of different products



unit-price (price per liter) paid on any shopping trip, i.e. $p_{t\mathbf{b}}$ divided by the number of units (liters) in **b**, on a constant, the number of units in **b**, and its square. In general, purchases of multiple units of CSDs by both household sizes are associated to lower prices per unit and quantity discounts are observed to decrease with each additional liter purchased. Moreover, multi-person households are observed, on average, to pay around 5 cents less than single-person households for any purchased liter of CSDs (the coefficient ΔConstant). These patterns are in line with the evidence by O'Connell and Smith (2020) on quantity discounts for soft drinks in the UK.

Collectively, the above descriptive evidence suggests that multi-person households tend to purchase more units of CSDs on any shopping trip and, because of quantity discounts, that they also tend to pay lower prices per liter (approximately 5 cents less per liter) than single-person households. While this supports the idea that firms may use quantity discounts as a screening device to price

Table 2: Descriptive Evidence on Quantity Discounts

	Price per Unit ($ per liter)
Single-person Households	
Constant	0.730 (0.009)
Num. Units	−0.021 (0.000)
(Num. Units)^2	0.0003 (0.0000)
Multi-person Households	
ΔConstant	−0.049 (0.002)
Num. Units	−0.012 (0.000)
(Num. Units)^2	0.0001 (0.000)
Product fixed effects	yes
Store fixed effects	yes
Time fixed effects	yes
Num. of Obs.	411,618
R^2	0.185

discriminate between households of different sizes (Maskin and Riley, 1984; Mussa and Rosen, 1978), it also highlights the imperfection of such mechanism, in that both multi-person and single-person households are observed to purchase multiple units of CSDs on any shopping trip (even though to a different extent). This and the oligopolistic nature of the industry complicate any a priori welfare assessment of the observed quantity discounts. In what follows, we rely on the PONL model to estimate the demand parameters of the different household sizes $hs \in \{\text{single}, \text{multi}\}$ and investigate some of the welfare implications of the observed quantity discounts in relation to linear pricing.

6.2 Model Specification

In this section, we specify our empirical PONL model of demand for multiple units of CSDs. We rely on the general notation introduced in section 2. The average utility of a household of size $hs \in$ {single, multi} in market t from purchasing a single unit of product j is:

$$\delta_{tj}^{hs} = \delta_j^{hs} - \alpha^{hs} p_{tj} + \delta_{\text{store}(t)} + \delta_{\text{time}(t)} + \xi_{tj}^{hs}, \tag{23}$$

where δ_j^{hs} is a household size and product-specific intercept, α^{hs} is a household size-specific price coefficient, $\delta_{\text{store}(t)}$ is a store fixed effect, $\delta_{\text{time}(t)}$ is a time (four weeks) fixed effect, and ξ_{tj}^{hs} is a residual observed by all economic agents (households and producers) but unobserved by the econometrician. The household size-specific nesting parameter for nest j is:

$$\lambda_j^{hs} = \lambda_{\text{Producer}(j)}^{hs}, \tag{24}$$

where Producer(j) $\in$ {Coca-Cola/PepsiCo, Dr.Pepper/Others} depending on the producer of product j. Following the discussion in section 2.1, each allocation parameter $\omega_{\mathbf{b}j}$ is equal to the reciprocal of the number of nests $\mathbf{b}$ belongs to:

$$\omega_{\mathbf{b}j} = \frac{1}{\sum_{j'=1}^{J} \mathbf{1}_{\mathbf{b}\in\mathbf{N}_{j'}}}$$

if $j \in \mathbf{b}$ and zero otherwise.

We use Hausman-type instruments (Hausman, 1996; Nevo, 2001) for the endogenous price p_{tj} and within-nest market share $s^{hs}_{t(j|j)}$. Remember that our markets are located in two cities, Pittsfield and Eau Claire. For the markets located in Pittsfield, we use the price of the same product j in the same retailer and four-week period (i.e., same time(t)) but as observed in Boston, the prices of products $r \neq j$ sold by the same retailer of j in the same time(t) but in Boston, the prices of products $k \neq j$ by the same producer of j as observed in the same time(t) but in Boston, and interactions of these. Similarly, for the markets located in Eau Claire, we use the same instruments as for Pittsfield but on the basis of the observed prices from Milwaukee.[37]

As mentioned in section 2.2, the demand synergy parameter $\Gamma^{hs}_{t\mathbf{b}}$ captures—among many other things—any indirect utility deviation due to nonlinear price $p_{t\mathbf{b}}$ relative to linear price $\sum_{j\in\mathbf{b}} p_{tj}$. To capture this, we decompose $\Gamma^{hs}_{t\mathbf{b}}$ as:

$$\Gamma^{hs}_{t\mathbf{b}} = -\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right) + \gamma^{hs}_{t\mathbf{b}}, \tag{25}$$

where $-\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$ isolates the part due to quantity discounts while $\gamma^{hs}_{t\mathbf{b}}$ captures every other potential source of synergy among the products in $\mathbf{b}$ (e.g., preference for variety or transportation costs). To further investigate the empirical determinants of demand for bundles and quantity discounts, given the thousands of estimated demand synergy parameters, we perform a second-step regression of $\hat{\gamma}^{hs}_{t\mathbf{b}} = \hat{\Gamma}^{hs}_{t\mathbf{b}} + \hat{\alpha}^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$ on observed characteristics and fixed effects:

$$\gamma^{hs}_{t\mathbf{b}} = \gamma^{hs}_{|\mathbf{b}|} + X_{\mathbf{b}}\gamma^{hs} + \gamma_{\text{store}(t)} + \gamma_{\text{time}(t)} + \eta^{hs}_{t\mathbf{b}}, \tag{26}$$

where $\gamma^{hs}_{|\mathbf{b}|}$ is a household size hs and bundle size $|\mathbf{b}|$-specific intercept, $X_{\mathbf{b}}$ is a vector of $\mathbf{b}$-specific characteristics, γ^{hs} is a household size-specific vector of parameters, $\gamma_{\text{store}(t)}$ is a store fixed effect, $\gamma_{\text{time}(t)}$ is a time (four weeks) fixed effect, and $\eta^{hs}_{t\mathbf{b}}$ is a residual term. As we illustrate next, interpreting the estimated demand synergies and the results of this second-step regression requires some care due to the (otherwise standard) normalization of the indirect utility of the outside option.

Normalization and Interpretation of Demand Synergy Parameters. Denote by δ^{hs}_{t0} the indirect utility of households size hs from choosing the outside option in market t. As is well known,

[37] Some retailers from Eau Claire are not present in Milwaukee. In these cases, we use the price of j, the prices of $k \neq j$, and the prices of $r \neq j$ by the same *producer* of j (and their interactions) as observed in Milwaukee.

the normalization $\delta_{t0}^{hs} = 0$ consists in subtracting δ_{t0}^{hs} from each $\delta_{t\mathbf{b}}^{hs}$, $\mathbf{b} \in \mathbf{C}_1$. As a result, the identified indirect utility of household size hs from purchasing bundle $\mathbf{b}$ corresponds to $\tilde{\delta}_{t\mathbf{b}}^{hs} = \delta_{t\mathbf{b}}^{hs} - \delta_{t0}^{hs}$ and, in turn, the identified demand synergies to:

$$\begin{aligned}\tilde{\Gamma}_{t\mathbf{b}}^{hs} &= \tilde{\delta}_{t\mathbf{b}}^{hs} - \sum_{j \in \mathbf{b}} \tilde{\delta}_{tj}^{hs} \\ &= (\delta_{t\mathbf{b}}^{hs} - \delta_{t0}^{hs}) - \sum_{j \in \mathbf{b}} \left[\delta_{tj}^{hs} - \delta_{t0}^{hs}\right] \\ &= \Gamma_{t\mathbf{b}}^{hs} + (|\mathbf{b}| - 1)\delta_{t0}^{hs}.\end{aligned} \tag{27}$$

Substituting this in (25) and (26), we obtain:

$$\begin{aligned}\tilde{\Gamma}_{t\mathbf{b}}^{hs} &= -\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j \in \mathbf{b}} p_{tj}\right) + \tilde{\gamma}_{t\mathbf{b}}^{hs}, \\ \tilde{\gamma}_{t\mathbf{b}}^{hs} &= \sum_{k=2}^{|\mathbf{b}|}\left(\gamma_k^{hs} - \gamma_{k-1}^{hs} + \delta_0^{hs}\right) + X_{\mathbf{b}}\gamma^{hs} + \gamma_{\text{store}(t)} + \gamma_{\text{time}(t)} + \tilde{\eta}_{t\mathbf{b}}^{hs},\end{aligned} \tag{28}$$

where $\tilde{\eta}_{t\mathbf{b}}^{hs} = \eta_{t\mathbf{b}}^{hs} + (|\mathbf{b}| - 1)(\delta_{t0}^{hs} - \delta_0^{hs})$, δ_0^{hs} is the average of δ_{t0}^{hs} over markets, and $\gamma_1^{hs} = 0$. One can then identify $\gamma_k^{hs} - \gamma_{k-1}^{hs} + \delta_0^{hs}$ for $k = 2, ..., |\mathbf{b}|$, but cannot separately identify $\gamma_k^{hs} - \gamma_{k-1}^{hs}$ and δ_0^{hs} without further assumptions. While this may complicate the interpretation of the estimated demand synergies, in that they will be "shifted" by $(|\mathbf{b}| - 1)\delta_{t0}^{hs}$, all the objects of interest (e.g., demand elasticities, marginal costs, consumer surplus, etc.) necessary to perform our counterfactual simulations are only functions of $\tilde{\gamma}_{t\mathbf{b}}^{hs}$—rather than of its individual components—and thus identified.

6.3 Estimation Results

Table 3 reports the C2SLS estimates of (23) and (24) from our iterative procedure, which (on our standard desktops) achieves numerical convergence in less than two minutes with $16,874$ bundles (with an average of 123.8 bundles per market) and a total of $176,700$ parameters.[38] The three columns of Table 3 summarize estimation results for three specifications of (24). In column (i) we assume a common nesting parameter across products and household sizes $\lambda_j^{hs} = \lambda$, in column (ii) we allow for two nesting parameters $\lambda_j^{hs} = \lambda^{hs}$, $hs \in \{\text{single, multi}\}$, while in column (iii) we specify $\lambda_j^{hs} = \lambda_{\text{Producer}(j)}^{hs}$ with $hs \in \{\text{single, multi}\}$ and $\text{Producer}(j) \in \{\text{Coca-Cola/PepsiCo, Dr.Pepper/Others}\}$. Standard errors are computed using the asymptotic formula detailed in Appendix B.1.

Table 3 suggests that single-person households are less price sensitive than multi-person households ($\alpha^{\text{single}} < \alpha^{\text{multi}}$) but also that nesting parameters are almost the same across household sizes and close to one, suggesting that—after controlling for all the fixed effects and demand synergies—the within-nest correlation in unobserved preferences is not very large (i.e., $(1 - \lambda_j^{hs}) \approx 0.1$). Despite the unconstrained estimation, all nesting parameters lie between 0 and 1, as required by consistency with

[38] Using the criterion discussed in footnote 31, we achieve numerical convergence after 25 iterations.

Table 3: Demand Estimates, Price Coefficients and Nesting Parameters

	(i): $\lambda_j^{hs} = \lambda$	(ii): $\lambda_j^{hs} = \lambda^{hs}$	(iii): $\lambda_j^{hs} = \lambda^{hs}_{\text{Producer}(j)}$
Price coefficients			
α^{single}	0.7369	0.7338	0.7479
	(0.1079)	(0.1079)	(0.1094)
α^{multi}	0.9992	1.0020	1.0257
	(0.0991)	(0.0994)	(0.1014)
Nesting parameters			
λ	0.8990		
	(0.0238)		
λ^{single}		0.8898	
		(0.0281)	
λ^{multi}		0.9028	
		(0.0254)	
$\lambda^{\text{single}}_{\text{Coca-Cola/PepsiCo}}$			0.8778
			(0.0301)
$\lambda^{\text{single}}_{\text{Dr.Pepper/Others}}$			0.9095
			(0.0454)
$\lambda^{\text{multi}}_{\text{Coca-Cola/PepsiCo}}$			0.9216
			(0.0270)
$\lambda^{\text{multi}}_{\text{Dr.Pepper/Others}}$			0.8549
			(0.0382)
Control for δ_j^{hs}	Yes	Yes	Yes
Store fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Num. of Obs.	$12,433$	$12,433$	$12,433$

Notes: The Table reports C2SLS estimates of (23) and (24) from the iterative procedure described in section 4. Standard errors are computed using the asymptotic formula detailed in Appendix B.1.

utility maximizing behavior (Ben-Akiva and Bierlaire, 1999; Bierlaire, 2006; Wen and Koppelman, 2001). In what follows, we rely on the estimates from column (iii), Table 3, as our preferred model.[39]

Table 4 reports results for the second-step OLS regression (28) of the estimated $\tilde{\gamma}_{t\mathbf{b}}^{hs} = \tilde{\Gamma}_{t\mathbf{b}}^{hs} + \alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$, where both $\tilde{\Gamma}_{t\mathbf{b}}^{hs}$ and α^{hs} are replaced by the C2SLS estimates of the specification from column (iii), Table 3.[40] While Table 3 does not report the C2SLS estimates of $\tilde{\Gamma}_{t\mathbf{b}}^{hs}$, the top panel of Table 4 reports their average net of quantity discounts, $\tilde{\gamma}_{t\mathbf{b}}^{hs}$, for single-person (first column) and multi-person households (second column). Even though we cannot directly interpret these (see discussion about the normalization of δ_0^{hs} above), the fact that multi-person households have on average larger $\tilde{\gamma}_{t\mathbf{b}}^{hs}$ (57.1 versus 36.7) suggests that they have stronger preferences (or needs) for larger quantities of CSDs, in line with the descriptive evidence from Figure 3. Differently, the fact that both household sizes have large average $\tilde{\gamma}_{t\mathbf{b}}^{hs}$ is consistent with the large observed share of shopping trips with no purchase of CSDs, 76.29% from Table 1, which leads to a large value of δ_0^{hs}.

The regression results in Table 4 highlight that quantity discounts do not fully explain demand

[39] Relying instead on those from column (ii), Table 3, leads to the same conclusions.

[40] As discussed in footnote 35, the C2SLS estimator cannot pin down the demand synergy $\Gamma_{t\mathbf{b}}$ of bundles that include products that are never observed to be purchased in isolation as single units in market t. In our PONL model, we have $81,215$ of these demand synergies. Because these cannot be part of the second-step OLS regression, the estimation sample used in Table 4 only includes $82,808$ of the $164,023$ total demand synergies. As detailed in Appendix E, even though we cannot pin down these $81,215$ demand synergies, we still account for them when estimating price elasticities and marginal costs, and when simulating counterfactuals.

Table 4: Demand Estimates: Demand Synergy Parameters

	hs = single	hs = multi
Average (std. dev.) $\tilde{\gamma}^{hs}_{t\mathbf{b}}$	36.695 (39.920)	57.124 (56.394)
OLS Estimation Results		
#producers in $\mathbf{b} = 2$	−2.499 (0.531)	−1.239 (0.110)
#producers in $\mathbf{b} = 3$	−4.856 (2.153)	−1.461 (0.302)
#producers in $\mathbf{b} = 4$	−8.179 (9.909)	−4.590 (1.589)
#products in $\mathbf{b} = 2$	2.002 (0.430)	−0.432 (0.099)
#products in $\mathbf{b} = 3$	6.663 (1.161)	1.330 (0.167)
#products in $\mathbf{b} = 4$	7.923 (3.404)	2.827 (0.323)
#products in $\mathbf{b} = 5$	8.647 (9.558)	6.818 (0.671)
#products in $\mathbf{b} = 6$	−	5.503 (1.094)
#products in $\mathbf{b} = 7$	−	12.485 (1.866)
#products in $\mathbf{b} = 8$	−	20.118 (3.818)
#products in $\mathbf{b} = 9$	−	45.993 (5.425)
Control for $\gamma^{hs}_k - \gamma^{hs}_{k-1} + \delta^{hs}_0$	yes	
Store fixed effects	yes	
Time fixed effects	yes	
Num. of Obs.	82,808	
R^2	0.972	

Notes: The Table reports results for the second-step OLS regression (28) of the demand synergy parameters as obtained from the first-step C2SLS estimates from column (iii), Table 3. "−" denotes that bundles with the corresponding characteristics for the given household size are not observed in the data and thus not included in the regression. Standard errors are computed using the basic OLS asymptotic formula.

synergies, and ultimately the purchase of multiple units of CSDs on any shopping trip. More precisely, $-\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$ explains between 23.19% and 91.12% of the total variance of $\Gamma^{hs}_{t\mathbf{b}}$.[41] Net of quantity discounts, households appear to enjoy purchases of wider varieties of CSDs but also to dislike mixing products by different producers. In other words, after controlling for quantity discounts, households seem to like purchasing different CSDs (e.g., one liter of Coke and one of Sprite better than two liters of Sprite) but within the variety offered by the same producer (e.g., Coke and Sprite better than Coke and 7Up). Moreover, a comparison between the two columns of Table 4 illustrates

[41] We compute the lower bound as one minus the ratio between the variance of $\tilde{\gamma}^{hs}_{t\mathbf{b}}$ and that of $\tilde{\Gamma}^{hs}_{t\mathbf{b}} = \Gamma^{hs}_{t\mathbf{b}} + (|\mathbf{b}|-1)\delta^{hs}_{t0}$, a measure of the extent to which $-\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$ explains variation in $\tilde{\Gamma}^{hs}_{t\mathbf{b}}$. This is a lower bound because we cannot tease out the part of variation in $\tilde{\Gamma}^{hs}_{t\mathbf{b}}$ explained by $(|\mathbf{b}|-1)\delta^{hs}_{t0}$. Similarly, we compute the upper bound as one minus the ratio between the variance of $X_{\mathbf{b}}\gamma^{hs} + \gamma_{\text{store}(t)} + \gamma_{\text{time}(t)}$ and that of $-\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right) + X_{\mathbf{b}}\gamma^{hs} + \gamma_{\text{store}(t)} + \gamma_{\text{time}(t)}$. This is an upper bound because $-\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right) + X_{\mathbf{b}}\gamma^{hs} + \gamma_{\text{store}(t)} + \gamma_{\text{time}(t)}$ only explains part of the variation in $\Gamma^{hs}_{t\mathbf{b}}$.

Table 5: Price Elasticities, by Household Size

Single-person households	$(p_{tj})_{j \in \mathbf{J}}$	$(p_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$
Single units		
$\sum_{j \in \mathbf{J}} \varepsilon_{tj}^{\text{single}}$	-1.0484 (0.146)	1.1851 (0.177)
Multiple units		
$\sum_{\mathbf{b} \in \mathbf{C}_2} \lvert\mathbf{b}\rvert \times \varepsilon_{t\mathbf{b}}^{\text{single}}$	0.0188 (0.003)	-5.0123 (0.685)
Multi-person households		
Single units		
$\sum_{j \in \mathbf{J}} \varepsilon_{tj}^{\text{multi}}$	-1.5072 (0.160)	3.1805 (0.452)
Multiple units		
$\sum_{\mathbf{b} \in \mathbf{C}_2} \lvert\mathbf{b}\rvert \times \varepsilon_{t\mathbf{b}}^{\text{multi}}$	0.0193 (0.002)	-7.0317 (0.603)

Notes: The Table reports the median of each price elasticity across those markets in which the two collections of bundles are observed to be purchased by both household sizes. We derive the expressions used to compute these price elasticities in Appendix F.1. Standard errors are computed using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

that single-person households have stronger preferences for this type of within-producer variety.

6.4 Estimated Elasticities

Our main objective is to evaluate the welfare effects of the observed quantity discounts by simulating a counterfactual in which producers implement linear pricing (i.e., a constant unit-price for each product), essentially a ban on nonlinear pricing. As a way to summarize our estimation results and provide intuition for this counterfactual simulation, Table 5 reports price elasticities of demand computed on the basis of the C2SLS estimates from column (iii), Table 3. These capture percentage changes in demand for a *collection* of bundles (Table rows: all single units $\sum_{j \in \mathbf{J}} \varepsilon_{tj}^{hs}$ and all multiple units $\sum_{\mathbf{b} \in \mathbf{C}_2} |\mathbf{b}| \times \varepsilon_{t\mathbf{b}}^{hs}$, where $|\mathbf{b}|$ is the number of units (liters) in bundle $\mathbf{b}$) with respect to a 1% increase in a *group* of prices (Table columns: all prices of single units $(p_{tj})_{j \in \mathbf{J}}$ and all prices of multiple units $(p_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$).[42] We derive the expressions of these price elasticities in Appendix F.1, while Table 5 reports the median of each across those markets in which the two collections of bundles are observed to be purchased by both household sizes. Standard errors are computed using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

Table 5 focuses on price elasticities with respect to two groups of prices, all prices of single units (first column) and all prices of multiple units (second column), as we expect these to be the most relevant when comparing quantity discounts to linear pricing. In the observed scenario of quantity discounts, the producer of each product j sets all quantity-specific prices of j: price p_{tj} for purchases

[42] We measure demand in liters of CSDs by weighing each bundle $\mathbf{b}$ by the number of units (liters), i.e. $|\mathbf{b}|$, it includes. In the context of demand for bundles, where each "alternative" $\mathbf{b}$ may correspond to different quantities, we find this measure more interpretable than the unweighted purchase probabilities.

of a single unit of j, $p_{t(j,j)}$ for purchases of two units of j, and so on, $p_{t(j,...,j)}$ for purchases of any number of units of j sold as a package. Then, the price of any bundle $\mathbf{b} \neq (j, ..., j)$ that combines different products other than j, $p_{t\mathbf{b}}$, is given by the sum of the quantity-specific prices of each product in $\mathbf{b}$. In the counterfactual scenario of linear pricing, instead, the producer of each product j sets only price p_{tj}, the unit-price of j, while the price of any bundle $\mathbf{b}$ is simply given by the sum of the unit-prices of its components $p_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} p_{tj}$. With linear pricing, producers lose their ability to set any element of $(p_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$ separately from $(p_{tj})_{j \in \mathbf{J}}$ and can instead only choose $(p_{tj})_{j \in \mathbf{J}}$.[43]

Remember from Table 3 that single-person households are less price sensitive than multi-person households ($\alpha^{\text{single}} < \alpha^{\text{multi}}$). This directly implies the main patterns reported in Table 5: multi-person households appear to be at least as price elastic as single-person households given the observed quantity discounts, both in terms of own-price and of cross-price effects. For example, a +1% increase in $(p_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$ would lead to a decrease of −7.03% in the purchases of multiple units by multi-person households, but only of −5.01% in those by single-person households. Symmetrically, this same +1% increase in $(p_{t\mathbf{b}})_{\mathbf{b} \in \mathbf{C}_2}$ would also lead to a +3.18% increase in the purchases of single units by multi-person households, but only of +1.19% in those by single-person households.

6.5 Counterfactual Simulation: Linear Pricing

Here we present our results on the welfare changes of quantity discounts. To evaluate these, we first rely on the PONL estimates from column (iii), Table 3, and calculate producers' marginal costs.[44] As detailed in Appendix F.2, we do this under the assumption that the observed prices were generated according to an oligopolistic Betrand-Nash price-setting game of complete information that allows each product to have quantity-specific prices. Importantly, this model does not assume producers to offer quantity discounts, but rather allows for the possibility that they choose to do so (along with the possibility of offering linear or even convex prices, i.e. increasing with quantity). Second, given the PONL estimates from column (iii), Table 3, and assuming that producers' marginal costs are invariant to the pricing strategy, we compute a vector of counterfactual linear prices for each market (independently across markets) following the procedure described in Appendix F.3. Finally, given the observed prices under quantity discounts and the simulated linear prices, we compute the implied changes in purchased quantities, profits, and compensating variations following the steps detailed in Appendix F.4. Tables 6 and 7 summarize these results in terms of median changes across the same set of markets used in Table 5.[45] As for the price elasticities, standard errors are computed using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

[43] Linear pricing is a constrained version of quantity discounts in which producers cannot choose the prices of multiple units of the same product $(p_{t(j,...,j)})_{j \in \mathbf{J}}$ separately from the corresponding unit-prices, in that $(p_{t(j,...,j)} = |(j, ..., j)| \times p_{tj})_{j \in \mathbf{J}}$, where $|(j, ..., j)|$ is the number of units of j in bundle $\mathbf{b} = (j, ..., j)$.

[44] We allow marginal costs to differ both across products and across numbers of units for each product, e.g. two units of j may have a different marginal cost than twice the marginal cost of one unit of j.

[45] Table 9 in Appendix F.6 reports the corresponding percentage changes with respect to the observed scenario of quantity discounts.

Table 6: Counterfactual Linear Pricing: Changes in Price and Quantity

	Average	Single-person households	Multi-person households
Price Change (\$)			
Δp_{tj}	-0.433 (0.014)		
$\Delta p_{t\mathbf{b}}$	$+0.881$ (0.070)		
Quantity Change (L per household-year)	-29.958 (1.483)	-12.018 (0.735)	-34.516 (1.732)
Single units	$+0.453$ (0.019)	$+0.262$ (0.027)	$+0.503$ (0.021)
Multiple units	-30.679 (1.512)	-12.312 (0.737)	-35.216 (1.791)
Quantity Change (%)	-20.38% (1.40%)	-17.64% (1.45%)	-20.99% (1.45%)
Conditional on purchase	-11.60% (0.79%)	-9.46% (0.93%)	-11.61% (0.89%)
ΔProb. of purchasing	-8.22% (0.70%)	-7.64% (0.78%)	-8.79% (0.70%)

Notes: We report all the computational details of the above entries in Appendices F.2 (marginal costs), F.3 (counterfactual simulation), and F.4 (price and quantity changes). All entries are computed as medians over the same set of markets used in Table 5 to compute price elasticities. Standard errors are obtained using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

The top panel of Table 6 shows that, in general, linear pricing would lead to a decrease in the prices of single units (up to one liter) of -43.3 cents and to a simultaneous increase in the prices of multiple units of $+88.1$ cents. With respect to the observed scenario of quantity discounts, these price changes are substantial and correspond to a decrease of -31.53% and to an increase of $+14.79\%$, respectively (Table 9, Appendix F.6). Intuitively, these price changes are expected to make purchases of smaller quantities relatively more convenient for both household sizes, and the middle panel of Table 6 confirms this: yearly purchased quantities per household would decrease by -29.96 liters, obtained as the difference between a small increase in purchases of single units ($+0.45$ liters) and a large reduction in purchases of multiple units (-30.68 liters).[46] The bottom panel of Table 6 shows that these large reductions in purchased quantities (-20.38%) are motivated by both a substitution from purchases of multiple units toward purchases of single units (-11.6%) and by a decrease in the probability of purchasing CSDs altogether (-8.22%).

Despite the generalized reduction in purchased quantities, as expected, linear pricing would induce heterogeneous responses in households of different sizes. To interpret these, one should bear in mind the purchasing patterns under quantity discounts documented in Figure 3 and the price elasticities in Table 5. While in relative terms multi-person households would reduce their purchased quantities only around 3 percentage points more than single-person households (-20.99% versus -17.64%, bottom panel, Table 6), the reductions in liters of CSDs purchased per year would look very different between household sizes (middle panel, Table 6): multi-person households would decrease their purchases

[46]To avoid problems with outliers, each of the entries of Tables 6, 7, and 9 is computed as a median across markets, so that the various decompositions of changes in quantities, profits, and compensating variations do not exactly add up to their totals. See Appendix F.4 for more details.

Table 7: Counterfactual Linear Pricing: Changes in Profit and Compensating Variation

	Average	Single-person households	Multi-person households
Profit Change ($ per household-year)	−7.827 (1.042)	−3.490 (0.521)	−8.922 (1.043)
Single units	+0.021 (0.017)	−0.051 (0.018)	+0.044 (0.020)
Multiple units	−8.018 (1.035)	−3.542 (0.534)	−9.148 (1.037)
CV ($ per household-year)	+3.556 (0.278)	+1.515 (0.234)	+3.856 (0.306)
Single units	−0.417 (0.015)	−0.371 (0.014)	−0.408 (0.016)
Multiple units	+3.974 (0.275)	+1.931 (0.244)	+4.410 (0.284)
CV/Expenditure (% per household-year)	+2.65% (0.12%)	+2.87% (0.36%)	+2.54% (0.11%)

Notes: CV denotes compensating variation. We report all the computational details of the above entries in Appendices F.2 (marginal costs), F.3 (counterfactual simulation), and F.4 (profit changes, CV, and CV/expenditure). All entries are computed as medians over the same set of markets used in Table 5 to compute price elasticities. Standard errors are obtained using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

by −34.52 liters per year, almost three times more than single-person households (−12.02 liters per year). The vast majority of this difference stems from the substantially larger reduction in purchases of multiple units by multi-person relative to single-person households (−35.22 versus −12.31 liters per year). This can be explained by noting that multi-person households both have higher price elasticity of demand for multiple units than single-person households (−7.03% versus −5.01%, Table 5) and purchase multiple units in far greater amounts under quantity discounts (panel c, Figure 3).

The top panel of Table 7 (Table 9, Appendix F.6, for the percentage changes) illustrates that this striking reduction in purchased quantities of −20.38% would cause a decrease in yearly profit per household of around −7.83$ (−20.53%), obtained as the difference between a very small per household-year profit increase from purchases of single units (+2 cents) and a very large per household-year profit reduction from purchases of multiple units (−8.02$). In line with the heterogeneous quantity changes reported in Table 6, producers would lose more than double yearly profit on multi-person households than on single-person households (−8.92$, or −21.27%, versus −3.49$, or −15.48%), simply losing more on those households whose purchased quantities would drop more sharply.

The middle and bottom panels of Table 7 show that a compensation of +3.56$ per household-year would be necessary for households to remain indifferent between quantity discounts and linear pricing, corresponding to 2.65% of their yearly expenditure on CSDs with quantity discounts. In line with the results from Table 6 and economic intuition, the compensating variation associated to linear pricing would vary between household sizes: while being generally small relative to yearly expenditure for all households, multi-person households would require more than double the compensation of single-person households: +3.86$ per household-year (2.54% of expenditure) as opposed to +1.52$ (+2.87% of expenditure). As we illustrate in Appendix F.5, these compensating variations can intuitively be understood in terms of the relative weights that households of different sizes place on the price changes.

Because multi-person households care relatively more about larger quantities and these would become more expensive, they would tend to lose more by linear pricing.

7 Concluding Discussion

From the above counterfactual simulation results, we can draw some important conclusions about quantity discounts and linear pricing. A first conclusion is that quantity discounts seem to be profitable for producers of CSDs in the USA. Despite the imperfect screening and the multi-product oligopolistic nature of the industry, this is in line with the standard textbook single-product monopoly model of quantity discounts with two types of consumers (Varian, 1992, pp. 244-248).

A second and perhaps more surprising conclusion is that, despite the substantial reduction in quantity purchased (−20.38%), consumer surplus would not reduce too sharply, with a compensating variation of +3.6$ per household-year (amounting to 2.7% of total expenditure on CSDs). This is the result of two countervailing forces: on the one hand, consumer surplus would decrease because of the contraction in purchases of larger quantities at relatively higher prices; on the other, however, it would increase because of the more frequent purchases of single units at relatively lower prices.[47]

These observations open up an important avenue for future research: a ban on quantity discounts could serve as a previously unexplored policy to limiting the consumption of CSDs and the intake of added sugar (Allcott et al., 2019; Bollinger et al., 2011; Dubois et al., 2020; O'Connell and Smith, 2020; Wang, 2015). Ricciuto et al. (2021) report that in the USA, in the period 2011-2012, 42.44% of the added sugar intake came from CSDs. Linear pricing would lead households to drastically reduce the purchased quantities of CSDs while only marginally reducing consumer surplus, potentially inducing large reductions in added sugar intake at the expense of a contraction in industry profit but none of the extra information (e.g., quantifying the marginal externality of added sugar) typically required to implement effective sugar taxes (Allcott et al., 2019; O'Connell and Smith, 2020).

Relying on additional nutrition label data and on the PONL estimates from column (iii), Table 3, we attempt a first step in this direction and simulate the reduction in purchased quantities of added sugar from CSDs implied by linear pricing. Our attempt here is subject to apparent limitations and is only intended to be illustrative of the potential for future investigations. We collect information on the amount of added sugar per liter for each of the 128 products included in our analysis from producers' and nutrition websites: as detailed in Table 10, Appendix F.6, 50% of the CSDs in our analysis have added sugar (the sugary CSDs), while the remaining 50% do not (the non-sugary CSDs).[48] Households purchase an average of 60L a year (51.2%) of sugary CSDs and 57.24L (48.8%) of non-sugary CSDs.

By following a procedure similar to that used in Table 6, we then compute the counterfactual

[47] As illustrated in Table 7, while the negative effect would slightly dominate the positive for all households, there would still be heterogeneity across household sizes. Multi-person households would substitute less away from the more expensive larger quantities toward the cheaper small quantities, facing larger losses in consumer surplus.

[48] We compute the amount of added sugar of each bundle by adding the amounts of added sugar per liter of its components.

Table 8: Counterfactual Linear Pricing: Changes in Added Sugar

	Amount per household-year	Percentage
Predicted added sugar change	$-1,753.6$g	-23.77%
	(166.875)	(2.40%)
Quantity change (average)	-29.958L	-20.38%
	(1.483)	(1.40%)
Sugary CSDs	-18.933L	-24.59%
	(2.070)	(2.71%)
Non-Sugary CSDs	-12.353L	-17.49%
	(1.041)	(1.86%)

Notes: The above changes are obtained in a similar manner to the counterfactual quantity changes in Table 6, see computational details in Appendix F.4. All entries are computed as medians over the same set of markets used in Table 5 to compute price elasticities.

change in purchased quantities of added sugar implied by linear pricing (see details in Appendix F.4). Table 8 summarizes these results in terms of median changes across the same set of markets used in Table 5. The top panel of Table 8 reports our simulated counterfactual: with linear pricing, households would reduce their yearly purchased quantities of added sugar from CSDs by -1.75kg (-23.77%). The bottom panel of Table 8 instead shows that linear pricing would induce a larger decrease in the yearly purchased quantities of sugary CSDs relative to non-sugary CSDs (-24.59% and -17.49%, respectively). This suggests that some of the decrease in purchased quantities of added sugar may be motivated by substitution from sugary to non-sugary CSDs.

Comparing these results to the existing literature, one notices that the decrease in purchased quantities of added sugar from CSDs implied by linear pricing may be of a similar order as that obtained by sugar taxes. For example, O'Connell and Smith (2020) find that an optimal tax in the UK would result in a reduction of -13.5% in the purchased quantities of sugary drinks and, in turn, a decrease of -28.4% in the purchased quantities of added sugar from soft drinks. Similarly, Dubois et al. (2020) find that a sugar tax of the form and size typically implemented in the UK and many US locations would lead to a reduction of around -21% in the purchased quantities of added sugar from soft drinks on-the-go. Seiler et al. (2021) document that a sugar tax introduced in Philadelphia led to a decrease of -16% in the purchased quantities of added sugar from soft drinks.[49]

Further research should investigate the many fundamental dimensions of comparison with sugar taxes we did not discuss, such as heterogeneity across income levels and in averted internalities. However, our results suggest that linear pricing *could* potentially serve as a policy to constrain the intake of added sugar from CSDs. Its main advantage over sugar taxes being the greater simplicity of implementation. Linear pricing can be obtained by banning quantity discounts and by enforcing that producers and retailers abide to this ban. Differently, the effective design and implementation of sugar taxes require not only demand and marginal cost estimates, but also information not always easy to obtain, for example on the externalities and the internalities associated to sugar intake (Allcott et al.,

[49] There are also studies that do not find significant effects of sugar taxes in the USA on the reduction of purchased quantities of added sugar from soft drinks, such as Bollinger and Sexton (2018); Rojas and Wang (2017); Wang (2015).

2019; O'Connell and Smith, 2020). In addition, sugar taxes require a more involved participation of governmental agencies to the market, especially for the collection and redistribution of tax revenue.

References

Abbe, E., M. Bierlaire, and T. Toledo (2007): "Normalization and correlation of cross-nested logit models," *Transportation Research Part B: Methodological*, 41(7), 795–808.

Adams, B., and K. R. Williams (2019): "Zone pricing in retail oligopoly," *American Economic Journal: Microeconomics*, 11(1), 124–56.

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

Aguirregabiria, V., and P. Mira (2002): "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 70(4), 1519–1543.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019): "Regressive sin taxes, with an application to the optimal soda tax," *The Quarterly Journal of Economics*, 134(3), 1557–1626.

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

——— (2020): "Latent complementarity in bundles models," *Available at SSRN 3257028*.

Allenby, G. M., T. S. Shively, S. Yang, and M. J. Garratt (2004): "A choice model for packaged goods: Dealing with discrete quantities and quantity discounts," *Marketing Science*, 23(1), 95–108.

Anderson, S. P., and L. Leruth (1993): "Why firms may prefer not to price discriminate via mixed bundling," *International Journal of Industrial Organization*, 11(1), 49–61.

Anderson, S. P., and R. Renault (2011): *Price discrimination*. Edward Elgar Publishing.

Armstrong, M. (2013): "A more general theory of commodity bundling," *Journal of Economic Theory*, 148(2), 448–472.

——— (2016): "Nonlinear pricing," *Annual Review of Economics*, 8, 583–614.

Aryal, G., and M. F. Gabrielli (2020): "An empirical analysis of competitive nonlinear pricing," *International Journal of Industrial Organization*, 68, 102538.

Ben-Akiva, M., and M. Bierlaire (1999): "Discrete choice methods and their applications to short term travel decisions," *Handbook of transportation science*, pp. 5–33.

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

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

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

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

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

Bhattacharya, D. (2018): "Empirical welfare analysis for discrete choice: Some general results," *Quantitative Economics*, 9(2), 571–615.

Bierlaire, M. (2006): "A theoretical analysis of the cross-nested logit model," *Annals of operations research*, 144(1), 287–300.

Bollinger, B., P. Leslie, and A. Sorensen (2011): "Calorie posting in chain restaurants," *American Economic Journal: Economic Policy*, 3(1), 91–128.

Bollinger, B., and S. Sexton (2018): "Local excise taxes, sticky prices, and spillovers: evidence from Berkeley's soda tax," *Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax (January 12, 2018)*.

Bonnet, C., and P. Dubois (2010): "Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance," *The RAND Journal of Economics*, 41(1), 139–164.

Bonnet, C., P. Dubois, S. B. Villas Boas, and D. Klapper (2013): "Empirical evidence on the role of nonlinear wholesale pricing and vertical restraints on cost pass-through," *Review of Economics and Statistics*, 95(2), 500–515.

Bresnahan, T. F., S. Stern, and M. Trajtenberg (1997): "Market segmentation and the sources of rents from innovation: Personal computers in the late 1980's," *The RAND Journal of Economics*.

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

Cardell, N. S. (1997): "Variance Components Structures for the Extreme Value and Logistic Distributions with Applications to Models of Heterogeneity," *Econometric Theory*, pp. 185–213.

Chan, T. Y. (2006): "Estimating a continuous hedonic-choice model with an application to demand for soft drinks," *The Rand journal of economics*, 37(2), 466–482.

Cho, S., and J. Rust (2010): "The flat rental puzzle," *The Review of Economic Studies*, 77(2), 560–594.

Chu, C. S., P. Leslie, and A. Sorensen (2011): "Bundle-size pricing as an approximation to mixed bundling," *The American Economic Review*, 101(1), 263–303.

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

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

Crawford, G., R. Griffith, and A. Iaria (forthcoming): "A Survey of Preference Estimation with Unobserved Choice Set Heterogeneity," .

Crawford, G. S., and M. Shum (2007): "Monopoly Quality Degradation in the Cable Television Industry," *The Journal of Law and Economics*, 50(1), 181–219.

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

Davis, P., and P. Schiraldi (2014): "The flexible coefficient multinomial logit (FC-MNL) model of demand for differentiated products," *The RAND Journal of Economics*, 45(1), 32–63.

DellaVigna, S., and M. Gentzkow (2019): "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 134(4), 2011–2084.

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

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," *Econometrica*, 80(5), 2231–2267.

Dubin, J., and D. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52(2), 345–362.

Dubois, P., R. Griffith, and M. O'Connell (2020): "How well targeted are soda taxes?," *American Economic Review*, 110(11), 3661–3704.

Ershov, D., J.-W. Laliberté, M. Marcoux, and S. Orr (2021): "Estimating Complementarity With Large Choice Sets: An Application to Mergers," *Working Paper*.

Florez-Acosta, J., and D. Herrera-Araujo (2020): "Multiproduct retailing and consumer shopping behavior: The role of shopping costs," *International Journal of Industrial Organization*, 68, 102560.

Fosgerau, M., J. Monardo, and A. de Palma (2021): "The Inverse Product Differentiation Logit Model," *Working Paper*.

Fox, J. T., and N. Lazzati (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.

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

Gandhi, A., Z. Lu, and X. Shi (2020): "Estimating demand for differentiated products with zeroes in market share data," *Available at SSRN 3503565*.

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

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2014): "Competition and ideological diversity: Historical evidence from us newspapers," *American Economic Review*, 104(10), 3073–3114.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019): "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 87(4), 1307–1340.

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

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

Guimaraes, P., and P. Portugal (2010): "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 10(4), 628–649.

Hausman, J. (1996): "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. Bresnahan, and R. Gordon. University of Chicago Press.

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

Hendel, I., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Stockpiling Behavior," *Econometrica*, 74(6), 1637–1673.

Hendel, I., and A. Nevo (2013): "Intertemporal price discrimination in storable goods markets," *American Economic Review*, 103(7), 2722–51.

Ho, K., J. Ho, and J. H. Mortimer (2012): "The use of full-line forcing contracts in the video rental industry," *The American Economic Review*, 102(2), 686–719.

Hospido, L. (2012): "Estimating Nonlinear Models with Multiple Fixed Effects: A Computational Note*," *Oxford Bulletin of Economics and Statistics*, 74(5), 760–775.

Huang, L., and Y. Liu (2017): "Health information and consumer learning in the bottled water market," *International Journal of Industrial Organization*, 55, 1–24.

Iaria, A., and A. Wang (2019): "Identification and Estimation of Demand for Bundles," *Available at SSRN 3458543*.

——— (2021): "A note on stochastic complementarity for the applied researcher," *Economics Letters*, 199, 109731.

Ivaldi, M., and D. Martimort (1994): "Competition Under Nonlinear Pricing," *Annales d'Economie et de Statistique*, 34.

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

Leslie, P. (2004): "Price Discrimination in Broadway Theatre," *Rand Journal of Economics*, pp. 520–541.

Levitt, S. D., J. A. List, S. Neckermann, and D. Nelson (2016): "Quantity discounts on a virtual good: The results of a massive pricing experiment at King Digital Entertainment," *Proceedings of the National Academy of Sciences*, 113(27), 7323–7328.

Lewbel, A., and L. Nesheim (2019): "Sparse demand systems: corners and complements," Discussion paper, cemmap working paper.

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

Luo, Y. (2018): "Bundling and nonlinear pricing in telecommunications," *Available at SSRN 3353876*.

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

Marzano, V., and A. Papola (2008): "On the covariance structure of the cross-nested logit model," *Transportation Research Part B: Methodological*, 42(2), 83–98.

Marzano, V., A. Papola, F. Simonelli, and R. Vitillo (2013): "A practically tractable expression of the covariances of the Cross-Nested Logit model," *Transportation Research Part B: Methodological*, 57, 1–11.

Maskin, E., and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15(2), 171–196.

McFadden, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Residential Location*, ed. by A. K. et. al. North Holland Pub. Co.

McManus, B. (2007): "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee," *Rand Journal of Economics*, 38, 512–532.

McManus, B., A. Nevo, Z. Nolan, and J. W. Williams (2020): "Steering Incentives of Platforms: Evidence from the Telecommunications Industry," .

Molina, H. (2020): "Buyer alliances in vertically related markets," *Available at SSRN 3452497*.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–17.

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

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

O'Connell, M., and K. Smith (2020): "Corrective tax design in oligopoly," *Working Paper*.

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

Reimers, I., and J. Waldfogel (2021): "Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings," *American Economic Review*, 111(6), 1944–71.

Ricciuto, L., V. L. Fulgoni III, P. C. Gaine, M. O. Scott, and L. DiFrancesco (2021): "Sources of Added Sugars Intake Among the US Population: Analysis by Selected Sociodemographic Factors Using the National Health and Nutrition Examination Survey 2011–18," *Frontiers in nutrition*, 8, 316.

Robert, C., and G. Casella (2013): *Monte Carlo statistical methods*. Springer Science & Business Media.

Rojas, C., and E. Y. Wang (2017): "Do taxes for soda and sugary drinks work? Scanner data evidence from Berkeley and Washington," *Scanner Data Evidence from Berkeley and Washington (September 23, 2017).*

Ruiz, F. J., S. Athey, and D. M. Blei (2020): "Shopper: A probabilistic model of consumer choice with substitutes and complements," *Annals of Applied Statistics*, 14(1), 1–27.

Rust, J. (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, pp. 999–1033.

Seiler, S., A. Tuchman, and S. Yao (2021): "The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects," *Journal of Marketing Research*, 58(1), 22–49.

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

Shiller, B., and J. Waldfogel (2011): "Music for a Song: An Empirical Look at Uniform Song Pricing and Its Alternatives," *Journal of Industrial Economics.*

Smith, H. (2006): "Store characteristics in retail oligopoly," *The RAND Journal of Economics*, 37(2), 416–430.

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

Song, M., S. Nicholson, and C. Lucarelli (2017): "Mergers with interfirm bundling: a case of pharmaceutical cocktails," *The RAND Journal of Economics*, 48(3), 810–834.

Su, C.-L., and K. L. Judd (2012): "Constrained optimization approaches to estimation of structural models," *Econometrica*, 80(5), 2213–2230.

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

Varian, H. R. (1989): "Price discrimination," *Handbook of industrial organization*, 1, 597–654.

——— (1992): "Microeconomic analysis," *3rd. ed. ed. New York, London: WW Norton & Company.*

Wang, A. (2019): "A BLP Demand Model of Product-Level Market Shares with Complementarity," *Working Paper.*

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.

Wen, C.-H., and F. S. Koppelman (2001): "The generalized nested logit model," *Transportation Research Part B: Methodological*, 35(7), 627–641.

Online Appendix for:

"An Empirical Model of Quantity Discounts with Large Choice Sets"

Alessandro Iaria and Ao Wang[1]

October, 2021

Appendix

A Proofs

A.1 Proof of Proposition 1

We rely on the following assumption for Proposition 1.

Assumption 1.

(i) For any $j = 1, ..., J$, $\mathbf{N}_j$ includes at least another bundle in addition to a single unit of j.

(ii) The support of $(p_t, X_t, (\mathscr{s}_{tj})_{j=1}^J)$ contains an open subset, where $p_t = (p_{tj})_{j=1}^J$ and $X_t = (x_{tj})_{j=1}^J$. Moreover, the support of $\mathscr{s}_t = (\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C_1}}$ is $\{(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C_1}} : \sum_{\mathbf{b}\in\mathbf{C}_1} \mathscr{s}_{t\mathbf{b}} < 1, \mathscr{s}_{t\mathbf{b}} > 0, \mathbf{b} \in \mathbf{C}_1\}$

(iii) $(p_t, X_t, \mathscr{s}_t)$ are complete for Z_t.

Assumption 1(i) is necessary for λ_j to be identifiable: if $\mathbf{N}_j$ only included j, then PONL model (7) would not depend on λ_j. Assumption 1(ii) requires a local support condition on $(p_t, X_t, (\mathscr{s}_{tj})_{j=1}^J)$ and a standard large support condition on $\mathscr{s}_t$. Assumption 1(iii) is a standard completeness condition in the literature on identification of demand using instrumental variables (Berry and Haile, 2014).

Define $\pi_t = (\pi_{tj})_{j=1}^J$. We first prove the following Lemma.

Lemma 1 (Uniqueness of Demand Synergies)**.** *Given $(\mathscr{s}_t, \lambda, \pi_t)$, $\Gamma_{t\mathbf{b}}$ is uniquely determined by* (16)*.*

Proof. Note that the left-hand side of (16) is increasing in $\Gamma_{t\mathbf{b}}$ while the right-hand side is decreasing in $\Gamma_{t\mathbf{b}}$. As $\Gamma_{t\mathbf{b}}$ increases from $-\infty$ to $+\infty$, the left- and the right-hand sides will cross only once. Consequently, given λ and π_t, (16) has a unique solution. □

[1]University of Bristol and CEPR (alessandro.iaria@bristol.ac.uk) and University of Warwick (ao.wang@warwick.ac.uk).

Due to Lemma 1, denote the unique solution to (16) by $\Gamma_{t\mathbf{b}} = \Gamma_{\mathbf{b}}(\lambda, \pi_t; \mathcal{s}_t)$. For any given λ and each $t \in \mathbf{T}$, (17) defines a system of J equations in $\pi_t \in \mathbb{R}_+^J$: for $j = 1, ..., J$,

$$\pi_{tj} = \left[\sum_{\mathbf{b}' \in \mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j} \exp\left(\Gamma_{\mathbf{b}'}(\lambda, \pi_t; \mathcal{s}_t)/\lambda_j\right) \prod_{r \in \mathbf{b}'} \left[\frac{\mathcal{s}_{tr}}{\mathcal{s}_{t0}}\right]^{\lambda_r/\lambda_j} \pi_{tr}^{(1-\lambda_r)/\lambda_j} \right]^{\lambda_j}. \tag{29}$$

Denote the true values of $(\alpha, \beta, \delta, \lambda)$ by $(\alpha^0, \beta^0, \delta^0, \lambda^0)$. Then, the true value $\pi_t = \pi_t^0$ is a solution to (29) when $\lambda = \lambda^0$. Denote $\pi_t^0 = \pi^0(\lambda^0; \mathcal{s}_t)$. Then, at $(\alpha^0, \beta^0, \delta^0, \lambda^0)$, for $j = 1, ..., J$,

$$\begin{aligned} &\ln \mathcal{s}_{tj} - \ln \mathcal{s}_{t0} = \delta_j^0 + x_{tj}\beta^0 - \alpha^0 p_{tj} + (1 - \lambda_j^0) \ln \left(\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_{tj}^0} \right) + \xi_{tj}, \\ &\mathbb{E}\left[g_j^0(\mathcal{s}_t, p_{tj}, x_{tj}; \alpha^0, \beta^0, \delta^0, \lambda^0) | Z_t = Z \right] = 0, \end{aligned} \tag{30}$$

where

$$g_j^0(\mathcal{s}_t, p_{tj}, x_{tj}; \alpha, \beta, \delta, \lambda) = \ln \mathcal{s}_{tj} - \ln \mathcal{s}_{t0} - \left(\delta_j + x_{tj}\beta - \alpha p_{tj} + (1 - \lambda_j) \ln \left(\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_j^0(\lambda; \mathcal{s}_t)} \right) \right).$$

Applying Assumption 1(iii), we identify $g_j^0 = g_j^0(\mathcal{s}_t, p_{tj}, x_{tj}; \alpha^0, \beta^0, \delta^0, \lambda^0)$ as a function of $(\mathcal{s}_t, p_{tj}, x_{tj})$ for each $j = 1, ..., J$. Then, due to Assumption 1(ii), we can use the derivatives of g_j^0 with respect to p_{tj} and x_{tj} to identify α^0 and β^0. Moreover, for $j = 1, ..., J$, by focusing on any market t such that $\mathcal{s}_{tj} \to 1$ and therefore $\ln \left(\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_j^0(\lambda^0;\mathcal{s}_t)} \right) = \ln \left(\mathcal{s}_{tj}/\mathcal{s}_t^j \right) \to 0$, we identify δ_j^0. As a result, we identify the quantities $A_{tj}^0 = (1 - \lambda_j^0) \ln \left(\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_j^0(\lambda^0;\mathcal{s}_t)} \right) = (1 - \lambda_j^0) \ln \left(\mathcal{s}_{tj}/\mathcal{s}_t^j \right)$ for $t \in \mathbf{T}$ and $j \in \mathbf{J}$. Using $\sum_{j=1}^J \pi_{tj}^0 + 1 = 1/\mathcal{s}_{t0}$, we obtain that for each $t \in \mathbf{T}$, $\lambda = \lambda^0$ satisfies:

$$\sum_{j=1}^J \frac{\mathcal{s}_{tj}}{1 - \mathcal{s}_{t0}} \exp\left(\frac{-A_{tj}^0}{1 - \lambda_j} \right) = 1 \tag{31}$$

or equivalently

$$\sum_{j=1}^J \frac{1}{1 - \mathcal{s}_{t0}} \left(\mathcal{s}_t^j \right)^{\frac{1-\lambda_j^0}{1-\lambda_j}} \mathcal{s}_{tj}^{\frac{\lambda_j^0 - \lambda_j}{1-\lambda_j}} = 1. \tag{32}$$

We now show that λ^0 is the only $\lambda \in \mathbb{R}^J$ that satisfies (32) and therefore identified. Because of Assumptions 1(i) and 1(ii), for each $j = 1, ..., J$, we can keep all $\mathcal{s}_{t\mathbf{b}}$, $\mathbf{b} \in \mathbf{N}_j$ and $\mathbf{b} \neq j$ constant and positive, while let $\mathcal{s}_{tj} \to 0$. Note that $\frac{1}{1-\mathcal{s}_{t0}} \left(\mathcal{s}_t^j \right)^{\frac{1-\lambda_j^0}{1-\lambda_j}}$ is always bounded away from zero and bounded from above for all $j = 1, ..., J$; in contrast, $\mathcal{s}_{tj}^{\frac{\lambda_j^0 - \lambda_j}{1-\lambda_j}}$ tends to $+\infty$ if $\lambda_j^0 < \lambda_j$. Then, for (32) to hold, we must have $\lambda_j^0 \geq \lambda_j$ for $j = 1, ..., J$. Note that given $\mathcal{s}_t$ and $\{A_{tj}^0\}_{j=1}^J$, the left-hand side of (31) is strictly increasing with respect to λ_j for $j = 1, ..., J$. Then, the only feasible λ that satisfies $\lambda_j^0 \geq \lambda_j$ and (31) is $\lambda = \lambda^0$. Finally, using Lemma 1, we identify all the $\Gamma_{t\mathbf{b}}$'s.

B Proof of Proposition 2

According to Lemma 1, given $\mathscr{s}_t$, λ, and π_t, $\Gamma_t = \Gamma(\lambda, \pi_t; \mathscr{s}_t)$ is uniquely determined. Plugging $\Gamma_t = \Gamma(\lambda, \pi_t; \mathscr{s}_t)$ in (17), we obtain: for $j = 1, ..., J$,

$$\pi_{tj} = \phi_j(\pi_t; \lambda, \mathscr{s}_t) = \left[\sum_{\mathbf{b}' \in \mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j} \exp\left(\Gamma_{\mathbf{b}'}(\lambda, \pi_t; \mathscr{s}_t)/\lambda_j\right) \prod_{r \in \mathbf{b}'} \left[\frac{\mathscr{s}_{tr}}{\mathscr{s}_{t0}}\right]^{\lambda_r/\lambda_j} \pi_{tr}^{(1-\lambda_r)/\lambda_j} \right]^{\lambda_j}. \tag{33}$$

Define $\Phi(\pi_t; \lambda, \mathscr{s}_t) = (\pi_{tj} - \phi_j(\pi_t; \lambda, \mathscr{s}_t))_{j=1}^{J}$.

Assumption 2. *There exist $a, b, \eta, M > 0$ such that*

$$\sup_{t \in \mathbf{T}, |\lambda - \lambda^0| \le a, |\pi_t - \pi_t^0| \le b} \left| Det \left(\frac{\partial \Phi(\pi_t; \lambda, \mathscr{s}_t)}{\partial \pi} \right) \right| > \eta, \tag{34}$$

and

$$\sup_{t \in \mathbf{T}, |\lambda - \lambda^0| \le a, |\pi_t - \pi_t^0| \le b} \left| Det \left(\frac{\partial \Phi(\pi_t; \lambda, \mathscr{s}_t)}{\partial \lambda} \right) \right| \le M, \tag{35}$$

where λ^0 and π_t^0 are the true values of λ and π_t, respectively. Moreover, $\frac{\partial \Phi(\pi_t; \lambda, \mathscr{s}_t)}{\partial \pi}$ is continuous at (π_t^0, λ^0), uniformly for $t \in \mathbf{T}$.

Regularity condition (34) is a rank condition of nonlinear constraints (33) with respect to π_t, uniformly for all $t \in \mathbf{T}$. This guarantees that each π_t can be expressed as a function of λ given $\mathscr{s}_t$ in a neighborhood of λ^0. Regularity condition (35) further ensures that this function from λ to π_t is uniformly bounded in the neighborhood of λ^0 uniformly for $t \in \mathbf{T}$.
We prove Proposition 2 in three steps.

Step 1: Uniqueness of π_t and Γ_t. Note that at $\lambda = \lambda^0$ and $\pi_t = \pi_t^0$, $\Phi(\pi_t^0; \lambda^0, \mathscr{s}_t) = 0$ for all $t \in \mathbf{T}$. Then, using Assumption 2 and applying the implicit function theorem, we can find $0 < d < a$ such that for any λ with $|\lambda - \lambda^0| < d$, there exists a unique π_t satisfying $|\pi_t - \pi_t^0| < b$ and $\Phi(\pi_t; \lambda, \mathscr{s}_t) = 0$ for all $t \in \mathbf{T}$.[50] Consequently, we can write $\pi_t = \pi(\lambda; \mathscr{s}_t)$ for λ with $|\lambda - \lambda^0| < d$ and all $t \in \mathbf{T}$. Then, Γ_t is also uniquely determined by λ and $\mathscr{s}_t$: $\Gamma_t = \Gamma(\lambda, \pi(\lambda; \mathscr{s}_t); \mathscr{s}_t)$.

Given the uniqueness of $\pi_t = \pi(\lambda; \mathscr{s}_t)$, we can then re-write (22) as:

$$\begin{aligned} \theta = (\delta, \beta, \alpha, 1 - \lambda) &= \psi_T(\theta) \\ &= \left(X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})X(\theta) \right)^{-1} X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})Y \end{aligned} \tag{36}$$

where $X(\theta) = (x_{tk}(\theta))_{t=1,...,T;k=1,...,2J+K+1} = \left((e_j)_{j=1}^{J}, x_t, -p_t, \left(\ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_j(\lambda; \mathscr{s}_t)} \right) \right)_{j=1}^{J} \right)_{t=1}^{T}$ and $Y = (y_t)_{t=1,...,T} = (\ln(\mathscr{s}_{t1}/\mathscr{s}_{t0}, ..., \mathscr{s}_{tJ}/\mathscr{s}_{t0}))_{t=1,...,T} \in \mathbb{R}^{JT \times 1}$.

[50] Because $\frac{\partial \Phi(\pi_t; \lambda, \mathscr{s}_t)}{\partial \pi}$ is continuous at (π_t^0, λ^0) uniformly for $t \in \mathbf{T}$ and because of the uniform lower bound η and upper bound M in Assumption 2, d does not depend on $t \in \mathbf{T}$.

Step 2: (Finite sample) Existence of a solution to (36). We now prove that when T is large enough, (36) has a solution in a fixed neighborhood of θ^0 (i.e., the neighborhood does not depend on T). Define

$$
\begin{aligned}
G_{XZ}(\theta) &= \left(\mathbb{E}\left[x_{tk}(\theta) z_{tk'}\right]\right)_{k=1,\ldots,2J+K+1, k'=1,\ldots,P} \in \mathbb{R}^{(2J+K+1)\times P},\\
G_{YZ} &= \left(\mathbb{E}\left[y_t z_{tk'}\right]\right)_{k'=1}^{P} \in \mathbb{R}^{P\times 1},\\
\psi(\theta) &= \left(G_{XZ}(\theta) G_{XZ}^{\mathrm{T}}(\theta)\right)^{-1} G_{XZ}(\theta) G_{YZ}.
\end{aligned}
$$

Assumption 3.

(i) There exists $r > 0$ and $d_0 \in (0, 1/2)$ such that for all $|\theta - \theta^0| \leq d_0$, and $v \in \mathbb{R}^{2J+K+1}$,

$$
r|v| \leq \left|\left(\frac{\psi(\theta)}{\partial\theta} - \mathbf{I}\right) v\right| \leq \frac{1}{r}|v|, \quad \left|\frac{\partial^2 \psi(\theta)}{\partial\theta^2}\right| \leq \frac{1}{r}.
$$

(ii)

$$
\sup_{|\theta-\theta^0|\leq d_0} \left\{ \left\{ \left| \frac{1}{T} \frac{\partial^k X^T(\theta) Z}{\partial\theta^k} - \frac{\partial^k G_{XZ}(\theta)}{\partial\theta^k} \right| \right\}_{k=0,1,2}, \left| \frac{Y^T Z}{T} - G_{YZ} \right| \right\} \xrightarrow{p} 0.
$$

Because of Assumption 3(ii), $\frac{\partial^k \phi_T(\theta)}{\partial\theta^k}$ converges uniformly to $\frac{\partial^k \phi(\theta)}{\partial\theta^k}$ in $|\theta - \theta^0| \leq d_0$ with probability one. Then, combining this with Assumption 3(i), we obtain that there exists $M_1 > 0$ such that

$$
\left|\left(\frac{\psi_T(\theta)}{\partial\theta} - \mathbf{I}\right)^{-1} v\right| \leq M_1 |v|, \quad \left|\left(\frac{\psi_T(\theta)}{\partial\theta} - \mathbf{I}\right)^{-1}\right| \times \left|\frac{\partial^2 \psi_T(\theta)}{\partial\theta^2}\right| \leq M_1 \tag{37}
$$

uniformly for $v \in \mathbb{R}^{2J+K+1}$ and $|\theta - \theta^0| \leq d_0$ with probability one as $T \to \infty$.

Now consider the following Newton-Raphson procedure:

$$
\begin{aligned}
\theta_0 &= \theta^0,\\
\theta_{k+1} &= -\left[\frac{\partial \psi_T(\theta_k)}{\partial\theta} - \mathbf{I}\right]^{-1} \left(\psi_T(\theta_k) - \theta_k\right) + \theta_k.
\end{aligned} \tag{38}
$$

Note that when $\theta = \theta^0$, we have $\pi_t = \pi_t^0$. Consequently, $\psi_T(\theta_0) = \psi_T(\theta^0)$ coincides with the (infeasible) 2SLS estimator obtained if we could observe $\mathfrak{z}_{t(j|j)}$, which we denote by θ_T^{2SLS}. Note that $\psi_T(\theta_0) = \theta_T^{\mathrm{2SLS}} \xrightarrow{p} \theta^0$ as $T \to \infty$.

Lemma 2. *Suppose that Assumptions 2-3 and* (37) *hold. In addition, $|\theta_T^{2SLS} - \theta^0| \leq \epsilon$, where $\epsilon > 0$ and $\epsilon \times M_1 < d_0/2$. Then, for any $k > 0$,$|\theta_k - \theta_{k-1}| \leq \left(\frac{d_0}{2}\right)^k$ and $|\psi_T(\theta_k) - \theta_k| \leq \epsilon \left(\frac{d_0}{2}\right)^k$.*

Proof. We prove Lemma 2 by induction. First, using Assumption 3, we have

$$
|\theta_1 - \theta_0| \leq M_1 \times |\psi_T(\theta_0) - \theta_0| = M_1 \times \left|\theta_T^{\mathrm{2SLS}} - \theta^0\right| \leq \frac{1}{2} d_0.
$$

Then, using the second-order Taylor expansion of $\psi_T((1-r)\theta_0 + r\theta_1) - ((1-r)\theta_0 + r\theta_1)$ around $r = 0$, the updating rule in (38), and Assumption 3, we obtain:

$$
\begin{aligned}
|\psi_T(\theta_1) - \theta_1| &= \left|\psi_T(\theta_0) - \theta_0 + \left[\frac{\partial \psi_T(\theta_0)}{\partial \theta} - \mathbf{I}\right](\theta_1 - \theta_0) + r_2(\theta_1 - \theta_0)\right| \\
&\leq M_1 \left|\psi_T(\theta_0) - \theta_0\right|^2 \\
&\leq M_1 \epsilon^2 \\
&\leq \frac{d_0 \epsilon}{2}.
\end{aligned}
$$

Suppose that the conclusions hold for k. We now prove that they hold for $k+1$. First, note that $|\theta_k - \theta_0| < d_0$. Then, using Assumption 3, we have

$$
|\theta_{k+1} - \theta_k| \leq M_1 \times |\psi_T(\theta_k) - \theta_k| = M_1 \times \epsilon \left(\frac{d_0}{2}\right)^k \leq \left(\frac{d_0}{2}\right)^{k+1}.
$$

Then, $|\theta_{k+1} - \theta_0| \leq |\theta_{k+1} - \theta_k| + |\theta_k - \theta_0| \leq \sum_{r=1}^{k+1} \left(\frac{d_0}{2}\right)^r$ and therefore $|\theta_{k+1} - \theta_0| \leq d_0$. Using Assumption 3, we obtain:

$$
\begin{aligned}
|\psi_T(\theta_{k+1}) - \theta_{k+1}| &= \left|\psi_T(\theta_k) - \theta_k + \left[\frac{\partial \psi_T(\theta_k)}{\partial \theta} - \mathbf{I}\right](\theta_{k+1} - \theta_k) + r_2(\theta_{k+1} - \theta_k)\right| \\
&\leq M_1 \left|\psi_T(\theta_k) - \theta_k\right|^2 \\
&\leq M_1 \epsilon^2 \left(\frac{d_0}{2}\right)^{2k} \\
&\leq \epsilon \left(\frac{d_0}{2}\right)^{2k+1} \\
&\leq \epsilon \left(\frac{d_0}{2}\right)^{k+1}.
\end{aligned}
$$

The proof is complete. □

Note that the event that (37) and $|\theta_T^{2SLS} - \theta^0| \leq \epsilon$ jointly hold occur with probability one as $T \to \infty$. Because $d_0 \in (0, 1/2)$, then Lemma 2 implies that with probability one: (1) θ_k converges to some θ^* such that $|\theta^* - \theta^0| \leq d_0$ and (2) $\psi_T(\theta^*) = \theta^*$, i.e. the existence of a solution to (36). Without loss of generality, define $\hat{\theta} = (\hat{\delta}, \hat{\beta}, \hat{\alpha}, 1 - \hat{\lambda}) = \theta^*$.

Step 3: Asymptotic properties of $\hat{\theta}$ and $(\hat{\pi}_t, \hat{\Gamma}_t)$. Because of the existence of a solution to (36), we can re-formulate $\hat{\theta}$ as an extremum estimator:

$$
\begin{aligned}
\hat{\theta} &= \underset{\theta : |\theta - \theta^0| \leq d_0}{\operatorname{argmin}} \; Q_T(\theta), \\
Q_T(\theta) &= \|\theta - \left(X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})X(\theta)\right)^{-1} X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})Y\|^2,
\end{aligned}
\tag{39}
$$

where $\|\cdot\|$ is the Euclidean distance. We rely on Theorem 2.1 of Newey and McFadden (1994) and verify the required conditions to show consistency. Define

$$Q(\theta) = \left\| \theta - \left(G_{XZ}(\theta) G_{XZ}^{\mathrm{T}}(\theta) \right)^{-1} G_{XZ}(\theta) G_{YZ} \right\|^2 .$$

Note that the true θ^0 satisfies $\theta^0 = \psi(\theta^0)$. Then, combining the implicit function theorem and Assumption (3)(i), we obtain the identification of θ^0 in a neighborhood of θ^0. This implies that $\theta = \theta^0$ is the unique minimizer of $Q(\theta) = 0$ in the compact set $\{\theta : |\theta - \theta^0| \leq d_0\}$. Moreover, due to the definition of $x_{tk}(\theta)$ and Assumption 3, $Q(\theta)$ is continuous. Finally, because of Assumption 3, $X^{\mathrm{T}}(\theta)Z/T \xrightarrow{p} G_{XZ}(\theta)$ uniformly for θ in $|\theta - \theta^0| \leq d_0$. Then, $Q_T(\theta) \xrightarrow{p} Q(\theta)$ uniformly for θ in $|\theta - \theta^0| \leq d_0$. The conditions of Theorem 2.1 by Newey and McFadden (1994) are verified and $\hat{\theta}$ is consistent.

To show the asymptotic normality of $\hat{\theta}$, we develop the first-order Taylor expansion of (36) at $\theta = \hat{\theta}$ around $\theta = \theta^0$:

$$\begin{aligned} 0 &= \hat{\theta} - \psi_T(\hat{\theta}) \\ &= \theta^0 - \psi_T(\theta^0) + \left[\mathbf{I} - \frac{\partial \psi_T(\tilde{\theta})}{\partial \theta} \right] \left(\hat{\theta} - \theta^0 \right) \\ &= \theta^0 - \theta_T^{2\mathrm{SLS}} + \left[\mathbf{I} - \frac{\partial \psi_T(\tilde{\theta})}{\partial \theta} \right] \left(\hat{\theta} - \theta^0 \right), \end{aligned}$$

where $\tilde{\theta}$ is a convex combination of θ^0 and $\hat{\theta}$. Then,

$$\sqrt{T} \left(\hat{\theta} - \theta^0 \right) = \left[\mathbf{I} - \frac{\partial \psi_T(\tilde{\theta})}{\partial \theta} \right]^{-1} \sqrt{T} \left(\theta_T^{2\mathrm{SLS}} - \theta^0 \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma V^{2\mathrm{SLS}} \Sigma^{\mathrm{T}}), \tag{40}$$

where $V^{2\mathrm{SLS}}$ is the asymptotic variance-covariance matrix of $\theta_T^{2\mathrm{SLS}}$ and

$$\Sigma = \left[\mathbf{I} - \frac{\partial \left[\left(G_{XZ}(\theta^0) G_{XZ}^{\mathrm{T}}(\theta^0) \right)^{-1} G_{XZ}(\theta^0) G_{YZ} \right]}{\partial \theta} \right]^{-1} . \tag{41}$$

The asymptotic normality of $\hat{\pi}_t$ and $\hat{\Gamma}_t$ follow from the uniqueness of π_t and Γ_t (as a function of $\hat{\theta}$ given $\jmath_t$) and the asymptotic normality of $\hat{\theta}$.

B.1 Inference

Here we describe how to conduct inference on θ and Γ_t, $t \in \mathbf{T}$, and objects that we derive from them in our counterfactual.

Inference on θ. We provide consistent estimators of $V^{2\text{SLS}}$ and Σ in (40). Given the consistency of $\hat{\theta}$, a plug-in estimator of $V^{2\text{SLS}}$ is:

$$\hat{V}^{2\text{SLS}} = \left(\hat{G}_{XZ}(\hat{\theta})\hat{G}_{XZ}^{\text{T}}(\hat{\theta})\right)^{-1} \hat{G}_{XZ}(\hat{\theta})\frac{Z^{\text{T}}\hat{\Omega}Z}{T}\hat{G}_{XZ}^{\text{T}}(\hat{\theta}) \left(\hat{G}_{XZ}(\hat{\theta})\hat{G}_{XZ}^{\text{T}}(\hat{\theta})\right)^{-1},$$

where $\hat{G}_{XZ}(\hat{\theta}) = \left(\frac{\sum_{t=1}^{T} x_{tk}(\hat{\theta})z_{tk'}}{T}\right)_{k=1,...,2J+K+1,k'=1,...,P} \in \mathbb{R}^{(2J+K+1)\times P}$ and $\hat{\Omega}$ is a consistent estimator of the variance-covariance matrix of ξ_t. Because of the definition of $x_{tk}(\hat{\theta})$, one can simply plug in $\hat{\pi}_t$, $t = 1, ..., T$.

Similarly, we can compute a plug-in estimator of Σ, denoted by $\hat{\Sigma}$. For this, it is sufficient to further compute $\frac{\partial G_{XZ}(\theta)}{\partial\theta}$ and G_{YZ}:

$$\begin{aligned}
&\frac{\partial\left[\left(G_{XZ}G_{XZ}^{\text{T}}\right)^{-1} G_{XZ}G_{YZ}\right]}{\partial\theta} \\
&= \left[\frac{\partial\left[\left(G_{XZ}G_{XZ}^{\text{T}}\right)^{-1}\right]}{\partial\theta}G_{XZ} + \left(G_{XZ}G_{XZ}^{\text{T}}\right)^{-1}\frac{\partial G_{XZ}}{\partial\theta}\right]G_{YZ} \\
&= \left(G_{XZ}G_{XZ}^{\text{T}}\right)^{-1}\left[\frac{\partial G_{XZ}}{\partial\theta} - \frac{\partial G_{XZ}}{\partial\theta}G_{XZ}^{\text{T}}(G_{XZ}G_{XZ}^{\text{T}})^{-1}G_{XZ} - G_{XZ}\frac{\partial G_{XZ}^{\text{T}}}{\partial\theta}(G_{XZ}G_{XZ}^{\text{T}})^{-1}G_{XZ}\right]G_{YZ}.
\end{aligned} \tag{42}$$

Then, we replace G_{XZ}, $\frac{\partial G_{XZ}(\theta)}{\partial\theta}$, and G_{YZ} by their finite-sample counterparts and $\theta = \hat{\theta}$ in (42) to obtain a consistent estimator of $\frac{\partial\left[\left(G_{XZ}G_{XZ}^{\text{T}}\right)^{-1}G_{XZ}G_{YZ}\right]}{\partial\theta}$. Finally, we plug this consistent estimator in (41) to obtain $\hat{\Sigma}$. When computing the empirical counterpart of $\frac{\partial G_{XZ}}{\partial\theta}$, we need to compute the derivative of π_t with respect to λ. To this end, we obtain their explicit formulae from (33).

Obtaining an explicit formula for $\hat{\Sigma}$ could be laborious in practice. We recommend instead a numerical alternative. The key is to compute the derivative $\frac{\partial\psi_T(\hat{\theta})}{\partial\theta}$, where $\psi_T(\hat{\theta})$ is defined as the 2SLS solution given $\pi_t = \pi(\hat{\lambda}; \jmath_t)$. Then, one can compute this derivative by the following central finite-difference formula:

$$\frac{\partial\psi_T(\hat{\theta})}{\partial\theta} = \frac{\psi_T(\hat{\theta}+h/2) - \psi_T(\hat{\theta}-h/2)}{h},$$

where h is small enough. Both $\psi_T(\hat{\theta}+h/2)$ and $\psi_T(\hat{\theta}-h/2)$ can be easily obtained using our proposed iterative procedure (see Appendix C for details). In practice, we iterate steps 1 and 3 at each iteration of the procedure (i.e., $\hat{\theta}$ is fixed). At the end of the procedure, we implement step 2 once more to obtain $\psi_T(\hat{\theta}+h/2)$ and $\psi_T(\hat{\theta}-h/2)$. We recommend this central finite-difference rather than forward (or backward) formulae because it is more robust to numerical errors caused by the iterative procedure.[51]

[51] The iterative procedure stops when the nonlinear system is approximately solved, giving rise to a very small numerical error. Intuitively, this numerical error is however orthogonal to the statistical error of the model. Moreover, it exists in both $\psi_T(\hat{\theta}+h/2)$ and $\psi_T(\hat{\theta}-h/2)$ computed using the iterative procedure. The proposed central finite-difference formula differences out this numerical error, achieving higher precision.

In our empirical application, we use $h = 10^{-6}$.

Inference on Γ_t. We recommend a parametric bootstrap method to conduct inference on Γ_t. For each $b = 1, ..., B$, we re-sample θ^b from asymptotic distribution of $\hat{\theta}$ in (40). Then, for each θ^b, we use the proposed iterative procedure to compute the corresponding Γ_t^b and construct its confidence interval using quantiles of the sample $\{\Gamma_t^b\}_{b=1}^B$. In the empirical application, we set $B = 200$.

Counterfactual Objects. Objects in the counterfactual are often functions of θ and Γ_t, $t \in \mathbf{T}$. We rely on the same parametric bootstrap method described above also to conduct inference on any counterfactual object.

C Details on the Iterative Estimation Procedure

C.1 Iteration 0: Choice of Starting Values

Similar to the practical implementation of any iterative procedure, one needs to set some starting values to launch the proposed algorithm. As intuition suggests, in extensive Monte Carlo simulations we noticed that the proposed iterative estimation procedure performs better (e.g., faster convergence and higher precision) when these starting values are closer to the true but unknown values of the parameters. The following three steps generate the starting values we found to perform best:

Step 0.1 For each (t, j) set $\pi_{tj}^{(0)} = \frac{\sum_{\mathbf{b}\in\mathbf{N}_j} \omega_{\mathbf{b}j} \mathcal{s}_{t\mathbf{b}}}{\mathcal{s}_{t0}}$, replacing each unobserved within-nest market share $\mathcal{s}_{t(\mathbf{b}|j)}$ by its corresponding (and observed) allocation parameter $\omega_{\mathbf{b}j}$.

Step 0.2. Given $\pi_t^{(0)}$, compute $\left(\delta^{(0)}, \lambda^{(0)}, \beta^{(0)}, \alpha^{(0)}\right)$ by 2SLS from the linear equations in (22), i.e. ignoring nonlinear constraints (16) and (17).

Step 0.3. Given $\pi_t^{(0)}$ and $\lambda^{(0)}$, for each $(t, \mathbf{b})$ independently compute $\Gamma_{t\mathbf{b}}^{(0)}$ by numerically solving constraint (16). This step can be executed in parallel for each $(t, \mathbf{b})$.

C.2 More Precise Formulation of the Algorithm

We provide some further mathematical detail on the formulae used in each step of the iterative estimation procedure. Given starting values $\left(\delta^{(0)}, \beta^{(0)}, \alpha^{(0)}, \lambda^{(0)}\right)$ and $\left(\pi_t^{(0)}, \Gamma_t^{(0)}\right)_{t=1}^T$, at each iteration k execute the following steps:

Step 1. (Direct update of π_{tj}) Given $\pi_t^{(k-1)}$, $\lambda^{(k-1)}$, and $\Gamma_t^{(k-1)}$, for each (t, j) compute $\pi_{tj}^{(k)}$ as a plug-in from the right-hand side of (17):

$$\pi_{tj}^{(k)} = \left[\sum_{\mathbf{b}'\in\mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j^{(k-1)}} \exp\left(\frac{\Gamma_{t\mathbf{b}'}^{(k-1)}}{\lambda_j^{(k-1)}} \right) \prod_{r\in\mathbf{b}'} \left[\frac{\mathcal{s}_{tr}}{\mathcal{s}_{t0}} \right]^{\lambda_r^{(k-1)}/\lambda_j^{(k-1)}} \left(\pi_{tr}^{(k-1)}\right)^{\left(1-\lambda_r^{(k-1)}\right)/\lambda_j^{(k-1)}} \right]^{\lambda_j^{(k-1)}} . \tag{43}$$

This step can be executed in parallel for each (t, j).

Step 2. Given $\pi_t^{(k)}$, compute $\left(\delta^{(k)}, \beta^{(k)}, \alpha^{(k)}, \lambda^{(k)}\right)$ by 2SLS as in Berry (1994) from the linear equations in (22), i.e. ignoring nonlinear equations (16) and (17).

Step 3. (One-step Newton-Raphson update of $\Gamma_{t\mathbf{b}}$) Given $\pi_t^{(k)}$, $\lambda^{(k)}$, and $\Gamma_t^{(k-1)}$, for each $(t, \mathbf{b})$ compute

$$G_{t\mathbf{b}}^{(k-1)} = \exp\left(\frac{\Gamma_{t\mathbf{b}}^{(k-1)}\left(1-\lambda_j^{(k)}\right)}{\lambda_j^{(k)}}\right) (\omega_{\mathbf{b}j})^{\frac{1}{\lambda_j^{(k)}}} \left(\pi_{tj}^{(k)}\right)^{1-\frac{1}{\lambda_j^{(k)}}} \prod\nolimits_{r\in\mathbf{b}} \left[s_{tr}/s_{t0}\right]^{\frac{\lambda_r^{(k)}\left(1-\lambda_j^{(k)}\right)}{\lambda_j^{(k)}}} \left(\pi_{tr}^{(k)}\right)^{\frac{\left(1-\lambda_r^{(k)}\right)\left(1-\lambda_j^{(k)}\right)}{\lambda_j^{(k)}}} \tag{44}$$

and then $\Gamma_{t\mathbf{b}}^{(k)}$ as the following plug-in:

$$\Gamma_{t\mathbf{b}}^{(k)} = \Gamma_{t\mathbf{b}}^{(k-1)} - \left[\Gamma_{t\mathbf{b}}^{(k-1)} - \left(\ln\left(\frac{s_{t\mathbf{b}}}{s_{t0}}\right) - \sum_{j\in\mathbf{b}}\left(\lambda_j^{(k)} \ln\left(\frac{s_{tj}}{s_{t0}}\right) + \left(1-\lambda_j^{(k)}\right)\ln\pi_{tj}^{(k)}\right)\right) + \ln\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k-1)}\right] \times \frac{\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k-1)}}{\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k-1)}/\lambda_j^{(k)}}.$$

This step can be executed in parallel for each $(t, \mathbf{b})$.

Step 4. If $k < \bar{K}$, move on to the next iteration $k+1$. If instead $k = \bar{K}$, exit the algorithm.

D Proof of Proposition 3

Our iterative estimation procedure is:

$$X^{(k)} = \left((e_j)_{j=1}^{J}, x_t, -p_t, \left(\ln\left(\frac{s_{tj}/s_{t0}}{\pi_{tj}^{(k)}}\right)\right)_{j=1}^{J}\right)_{t=1}^{T},$$

$$(\alpha^{(k+1)}, \beta^{(k+1)}, \delta^{(k+1)}, 1-\lambda^{(k+1)}) = \left(X^{(k)\mathrm{T}}(ZZ^{\mathrm{T}})X^{(k)}\right)^{-1}\left(X^{(k)\mathrm{T}}(ZZ^{\mathrm{T}})Y\right),$$

$$\Gamma_{t\mathbf{b}}^{(k+1)} = \Gamma_{t\mathbf{b}}^{(k)} - \left(\Gamma_{t\mathbf{b}}^{(k)} - \left(\ln[s_{t\mathbf{b}}/s_{t0}] - \sum_{j\in\mathbf{b}}\left(\lambda_j^{(k+1)}\ln[s_{tj}/s_{t0}] + \left(1-\lambda_j^{(k+1)}\right)\ln\pi_{tj}^{(k)}\right)\right) + \ln\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k)}\right)$$
$$\times \frac{\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k)}}{\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k)}/\lambda_j^{(k+1)}}$$

$$\pi_{tj}^{(k+1)} = \left[\sum_{\mathbf{b}'\in\mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j^{(k+1)}} \exp\left(\Gamma_{\mathbf{b}'}^{(k+1)}/\lambda_j^{(k+1)}\right) \prod_{r\in\mathbf{b}'}\left[\frac{s_{tr}}{s_{t0}}\right]^{\lambda_r^{(k+1)}/\lambda_j^{(k+1)}} \left[\pi_{tr}^{(k)}\right]^{(1-\lambda_r^{(k+1)})/\lambda_j^{(k+1)}}\right]^{\lambda_j^{(k+1)}},$$

where $G_{t\mathbf{b}}^{(k)}$ is defined in (44). Because $\pi_t^{(k)} \to \pi_t^*$ for all $t \in \mathbf{T}$, then $X^{(k)}$ as well as $(\alpha^{(k)}, \beta^{(k)}, \delta^{(k)}, 1-\lambda^{(k)})$ converge. Denote by $(\alpha^*, \beta^*, \delta^*, 1-\lambda^*)$ the limit of $(\alpha^{(k)}, \beta^{(k)}, \delta^{(k)}, 1-\lambda^{(k)})$. Because $\Gamma_{t\mathbf{b}}^{(k)} \to \Gamma_{t\mathbf{b}}^*$, then we obtain:

$$
\begin{aligned}
\Gamma^*_{t\mathbf{b}} = {} & \ln[\mathscr{s}_{t\mathbf{b}}/\mathscr{s}_{t0}] - \sum_{j\in\mathbf{b}} \left(\lambda^*_j \ln[\mathscr{s}_{tj}/\mathscr{s}_{t0}] + \left(1-\lambda^*_j\right)\ln \pi^*_{tj}\right) \\
& - \ln\left(\sum_{j=1}^{J} \exp\left(\frac{\Gamma^*_{t\mathbf{b}}(1-\lambda^*_j)}{\lambda^*_j}\right) (\omega_{\mathbf{b}j})^{\frac{1}{\lambda^*_j}} \left(\pi^*_{tj}\right)^{1-\frac{1}{\lambda^*_j}} \prod_{r\in\mathbf{b}} [\mathscr{s}_{tr}/\mathscr{s}_{t0}]^{\frac{\lambda^*_r(1-\lambda^*_j)}{\lambda^*_j}} \left(\pi^*_{tr}\right)^{\frac{(1-\lambda^*_r)(1-\lambda^*_j)}{\lambda^*_j}}\right)
\end{aligned}
$$

Consequently, at $(\alpha^*, \beta^*, \delta^*, 1-\lambda^*)$, $\Gamma^*_{t\mathbf{b}}$, and π^*_t, constraints (16) are satisfied. Similarly, constraints (17) are satisfied. Therefore, $(\delta^*, \beta^*, \alpha^*, \lambda^*)$ and $(\pi^*_t, \Gamma^*_t)_{t\in\mathbf{T}}$ satisfy (22).

E Dealing with Products with Undefined $\mathscr{s}_{tj}$

Without further assumptions, the C2SLS estimator cannot pin down the demand synergy $\Gamma_{t\mathbf{b}}$ of bundles that include products which are not observed to be purchased in isolation as single units in market t. This can be readily seen in equation (16): if product j' is not observed to be purchased in isolation as a single unit in market t, then $\mathscr{s}_{tj'}$ is not defined and, consequently, any $\Gamma_{t\mathbf{b}}$ corresponding to a $\mathbf{b}$ that includes j' will not be defined in market t. Whenever the incidence of bundles of this type is not very prominent, a simple solution is just be to exclude them from the analysis. However, when there are many of these bundles, excluding them may correspond to dropping a large share of purchases and may not be advisable. In this Appendix, we provide a practical solution to this problem that exploits all the available data (i.e., does not involve excluding these bundles from the analysis) and does not require any modification of the C2SLS estimator.

The main idea of the proposed approach is simple and consists of three steps. In the first step, we "separate away" from bundles any sub-bundle collecting those products whose purchase probability of a single unit $\mathscr{s}_{tj}$ is defined. In the second step, we implement the C2SLS estimator only on the products observed to be purchased as single units in isolation (i.e., with defined $\mathscr{s}_{tj}$) and the corresponding bundles and sub-bundles obtained in the first step. In the third step, we rely on the C2SLS estimates and the observed purchase probabilities to recover the average utility $\delta_{t\mathbf{b}}$ of those sub-bundles not used in the C2SLS estimation, i.e. those made of products whose purchase probability of a single unit $\mathscr{s}_{tj}$ is *not* defined. Given these, we then proceed to the computation of price elasticities, marginal costs, and counterfactual simulations as detailed in Appendix F.

Suppose the J' products in $\mathbf{J}'$, indexed by $j' = 1, ..., J'$, are only observed to be purchased as part of bundles, but not in isolation as single units. These products have undefined $\mathscr{s}_{tj'}$. In the first step, we partition any bundle $\mathbf{b}$ that includes at least one unit of any product in $\mathbf{J}'$ in at most $J'+1$ sub-bundles of the form $\mathbf{b} = \left((\mathbf{b}_{j'})_{j'=1}^{J'}, \mathbf{b}_{-\mathbf{J}'}\right)$, where each $\mathbf{b}_{j'} = (j', ..., j')$ collects all units of product j' in $\mathbf{b}$ and $\mathbf{b}_{-\mathbf{J}'}$ is the complement of $(\mathbf{b}_{j'})_{j'=1}^{J'}$.[52] To save on notation, we use the symbol $\mathbf{b}_{-\mathbf{J}'}$ also to

[52]For each bundle $\mathbf{b}$ and product $j' \in \mathbf{J}'$, $\mathbf{b}_{j'}$ could be empty if $\mathbf{b}$ does not include even one unit of j', $\mathbf{b}_{j'} = j'$ if $\mathbf{b}$ includes exactly one unit of j', $\mathbf{b}_{j'} = (j', j')$ if $\mathbf{b}$ includes two units of j', and so on. Because each of the J' sub-bundles $\mathbf{b}_{j'}$ can be empty, $\mathbf{b}$ will be partitioned in "up to" $J'+1$ non-empty sub-bundles.

refer to the original bundles $\mathbf{b}$ that do not include any purchase of products in $\mathbf{J}'$ and can directly be used in estimation. In the second step, we implement the C2SLS estimator on the $\mathbf{J}/\mathbf{J}'$ products and the bundles and sub-bundles denoted by $\mathbf{b}_{-\mathbf{J}'}$. Finally, in the third step, given the C2SLS estimates and the observed purchase probabilities, we recover the remaining average utilities $\delta_{t\mathbf{b}_{j'}}$, $j' \in \mathbf{J}'$. By re-writing the average utility of $\mathbf{b}_{j'}$ as in equation (13), $\delta_{t\mathbf{b}_{j'}} = \delta_{\mathbf{b}_{j'}} + x_{t\mathbf{b}_{j'}}\beta - \alpha p_{t\mathbf{b}_{j'}} + \xi_{t\mathbf{b}_{j'}}$, we can back out its remaining unknown component simply as:

$$\delta_{\mathbf{b}_{j'}} + \xi_{t\mathbf{b}_{j'}} = \ln(\mathscr{s}_{t\mathbf{b}_{j'}}) - \ln(\mathscr{s}_{t0}) - x_{t\mathbf{b}_{j'}}\beta + \alpha p_{t\mathbf{b}_{j'}} - (1-\lambda_j)\ln\left(\mathscr{s}_{t(\mathbf{b}_{j'}|j')}\right), \tag{45}$$

where $\mathscr{s}_{t(\mathbf{b}_{j'}|j')}$ is known because of the specific way we partitioned bundles in the first step: *(i)* $\mathbf{b}_{j'} = (j', ..., j')$ only belongs to nest $\mathbf{N}_{j'}$ and *(ii)* nest $\mathbf{N}_{j'}$ only includes bundles made of a single or multiple units of j' (i.e., it only includes sub-bundles $\mathbf{b}_{j'}$).

After having recovered $\delta_{\mathbf{b}_{j'}} + \xi_{t\mathbf{b}_{j'}}$ from (45) for all "problematic" sub-bundles $\mathbf{b}_{j'}$, $j' = 1, ..., J'$, we can proceed without further complications to computing price elasticities, marginal costs, and counterfactual simulations as detailed in Appendix F. The results presented in the empirical application in section 6 rely on this procedure. However, in unreported robustness checks, we repeated the empirical analysis by excluding all bundles $\mathbf{b}$ that include at least one unit of any product $j' \in \mathbf{J}'$ and—despite the smaller sample used—our estimates and counterfactual simulation results remain qualitatively similar.

F Empirical Application

F.1 Elasticities in Table 5

To simplify exposition, we drop the indexes of household size hs and of market t. Here we derive expressions for the demand elasticities we report, respectively, in the first and in the second column of Table 5: the percentage changes in the collective units for single-unit products in $\mathbf{J}$ and the multi-unit bundles in $\mathbf{C}_2$ due to a 1% increase in all prices of the single-unit products in $\mathbf{J}$ and that due to a 1% increase in all prices of multi-unit bundles in $\mathbf{C}_2$. We denote these elasticities by $E_{\mathbf{AB}}$ for $\mathbf{A}, \mathbf{B} \in \{\mathbf{J}, \mathbf{C}_2\}$.

$$E_{\mathbf{AB}} = \frac{\sum_{\mathbf{b}\in\mathbf{A}} |\mathbf{b}| \times \sum_{\mathbf{b}'\in\mathbf{B}} \frac{\partial \mathscr{s}_{\mathbf{b}}}{\partial p_{\mathbf{b}'}} p_{\mathbf{b}'}}{\sum_{\mathbf{b}\in\mathbf{A}} |\mathbf{b}| \times \mathscr{s}_{\mathbf{b}}} = \frac{\sum_{\mathbf{b}\in\mathbf{A}} |\mathbf{b}| \times \mathscr{s}_{\mathbf{b}} \sum_{\mathbf{b}'\in\mathbf{B}} \epsilon_{\mathbf{b}\mathbf{b}'}}{\sum_{\mathbf{b}\in\mathbf{A}} |\mathbf{b}| \times \mathscr{s}_{\mathbf{b}}}$$

where $|\mathbf{b}|$ is the number of units (liters) in bundle $\mathbf{b}$ and $\epsilon_{\mathbf{b}\mathbf{b}'}$ is the cross-price elasticity of $\mathbf{b}$ with respect to $p_{\mathbf{b}'}$:

$$\epsilon_{\mathbf{b}\mathbf{b}'} = -\frac{\alpha p_{\mathbf{b}'}}{\mathscr{s}_{\mathbf{b}}} \left[\sum_{j=1}^{J} \left[\left(1 - \frac{1}{\lambda_j}\right) \mathscr{s}_{\mathbf{b}|j} \times \mathscr{s}_{\mathbf{b}'|j} \times \mathscr{s}^j + \mathbf{1}_{\mathbf{b}=\mathbf{b}'} \frac{1}{\lambda_j} \mathscr{s}^j \times \mathscr{s}_{\mathbf{b}|j} \right] - \mathscr{s}_{\mathbf{b}'} \mathscr{s}_{\mathbf{b}} \right].$$

F.2 Computation of Marginal Costs

In the observed scenario, producers set single-unit prices for their products, e.g. p_j, as well as for bundles of multiple units of the same product, e.g. $p_{(j,\ldots,j)}$. Denote by $\mathbf{J}_1$ the set of single-unit products (where $\mathbf{J}_1 = \mathbf{J}$) and by $\mathbf{J}_2$ the set of bundles of multiple units of the same products, e.g. (j,j), (k,k), or (k,k,k). We rely on vector $m_{\mathbf{b}} \in \{0,1\}^{(|\mathbf{J}_1|+|\mathbf{J}_2|)}$, with $m_{\mathbf{b}\ell} \in \{0,1\}$ corresponding to element ℓ in $\mathbf{J}_1 \cup \mathbf{J}_2$, to describe the composition of bundle $\mathbf{b}$ in terms of elements of $\mathbf{J}_1 \cup \mathbf{J}_2$. For example, if $\mathbf{b} = (1,2,3,3,3)$, $\mathbf{J}_1 = \{1,2,3\}$, and $\mathbf{J}_2 = \{(1,1),(2,2),(3,3,3)\}$, then $m_{\mathbf{b}} = (1,1,0,0,0,1)$, with—for instance—second element $m_{\mathbf{b}2} = 1$ and fifth element $m_{\mathbf{b}(2,2)} = 0$.

To compute the producers' marginal costs given PONL estimates, we assume that the observed prices in the data were generated by an oligopolistic Bertrand-Nash price-setting game of complete information that allows each product to have quantity-specific prices. We allow the marginal costs to be specific to any product-quantity combination (e.g., could be cheaper to produce larger quantities) but assume that are not affected by the pricing scheme (will hold them constant in the counterfactual linear pricing). Denote by $\mathbf{O}$ the ownership matrix in the observed scenario in the data. This matrix is of dimension $(|\mathbf{J}_1| + |\mathbf{J}_2|) \times (|\mathbf{J}_1| + |\mathbf{J}_2|)$, and the element at position (k,ℓ), $o_{k,\ell} = 1$ if k and ℓ in $\in \mathbf{J}_1 \cup \mathbf{J}_2$ are owned by the same producer, or 0 otherwise. For example, $o_{1,2} = 1$ if products 1 and 2 are owned by the same producer, or 0 otherwise. Moreover, $o_{1,(1,1)} = 1$ because multiple units of the same product are still sold by the same producer.

Define $\mathbf{M} = (m_{\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1} \in \mathbb{R}^{C_1 \times (|\mathbf{J}_1|+|\mathbf{J}_2|)}$. Then, the first-order conditions (FOCs) of the oligopolistic Bertrand-Nash price-setting game in the observed scenario with quantity discounts can be written as:

$$F(p_{\mathbf{J}_1\cup\mathbf{J}_2}) = \left[\mathbf{O} \circ \left(\mathbf{M}^{\mathrm{T}} \frac{\partial \mathscr{s}_{\mathbf{C}_1}}{\partial p_{\mathbf{C}_1}} \mathbf{M}\right)\right] (p_{\mathbf{J}_1\cup\mathbf{J}_2} - c_{\mathbf{J}_1\cup\mathbf{J}_2}) + \mathbf{O}^{\mathrm{T}} \mathscr{s}_{\mathbf{C}_1} = 0_{(|\mathbf{J}_1|+|\mathbf{J}|_2)\times 1}, \tag{46}$$

where $\mathscr{s}_{\mathbf{C}_1}$ and $p_{\mathbf{C}_1}$ are the vectors of purchase probabilities and prices of all bundles in $\mathbf{C}_1$, and $p_{\mathbf{J}_1\cup\mathbf{J}_2}$ and $c_{\mathbf{J}_1\cup\mathbf{J}_2}$ are the vectors of prices and marginal costs of the single and multiple units of all products in $\mathbf{J}_1 \cup \mathbf{J}_2$. Importantly, the FOCs in (46) do not assume producers to offer quantity discounts, but rather allow for the possibility that they choose to do so (along with the possibility of offering linear or even convex prices, i.e., increasing with quantity). Note that each $\mathscr{s}_{\mathbf{b}}$ in $\mathscr{s}_{\mathbf{C}_1}$ is a weighted sum of the household size-specific purchase probabilities of $\mathbf{b}$ (i.e., our PONL estimates):

$$\mathscr{s}_{\mathbf{b}} = \sum_{hs} w_{hs} \mathscr{s}_{\mathbf{b}}^{hs},$$

and therefore,

$$\frac{\partial \mathscr{s}_{\mathbf{C}_1}}{\partial p_{\mathbf{C}_1}} = -\sum_{hs} w_{hs} \alpha^{hs} \frac{\partial \mathscr{s}_{\mathbf{C}_1}^{hs}}{\partial \delta_{\mathbf{C}_1}^{hs}},$$

where w_{hs} is the weight of household size hs in the population and $\delta_{\mathbf{C}_1}^{hs}$ is the vector of the average utilities of the bundles in $\mathbf{C}_1$ among the households of size hs. Then, we can back out the vector of marginal costs $c_{\mathbf{J}_1\cup\mathbf{J}_2}$ from FOCs (46):

$$c_{\mathbf{J}_1\cup\mathbf{J}_2} = p_{\mathbf{J}_1\cup\mathbf{J}_2} - \left[\mathbf{O}\circ\left(\mathbf{M}^{\mathrm{T}}\left(\sum_{hs=1}^{2} w_{hs}\alpha^{hs}\frac{\partial \mathcal{s}^{hs}_{\mathbf{C}_1}}{\partial \delta^{hs}_{\mathbf{C}_1}}\right)\mathbf{M}\right)\right]^{-1}\left(\mathbf{O}^{\mathrm{T}}\mathcal{s}_{\mathbf{C}_1}\right).$$

F.3 Counterfactual Simulation: Linear Pricing

To simulate the counterfactual scenario with linear pricing, we start from the setting in Appendix F.2 and rule out quantity-specific prices for every product j: $p_{(j,...,j)}$ for any $(j,...,j)\in\mathbf{J}_2$ equals $|(j,...,j)|$ times p_j, $j\in\mathbf{J}_1$. In practice, we do this by setting the term capturing quantity discount $-\alpha^{hs}\left(p_{\mathbf{b}}-\sum_{j\in\mathbf{b}}p_j\right) = 0$ in the estimated demand synergy (28), so that $\tilde{\Gamma}^{hs}_{\mathbf{b}} = \tilde{\gamma}^{hs}_{\mathbf{b}}$ for $hs\in\{\text{single},\text{multi}\}$ and $\mathbf{b}\in\mathbf{C}_2$, and by letting producers re-optimize with respect to $p_{\mathbf{J}_1} = (p_j)_{j\in\mathbf{J}_1}$ as described in what follows.

Define a matrix of dimension $(|\mathbf{J}_1|+|\mathbf{J}_2|)\times|\mathbf{J}_1|$, M_{12}, whose (ℓ,k) element is equal to the number of units of product $k\in\mathbf{J}_1$ in $\ell\in\mathbf{J}_1\cup\mathbf{J}_2$. For example, if $\ell=1$ and $k=1$, then the corresponding element in M_{12} is 1. If $\ell=(1,1)$ and $k=1$, then the corresponding element in M_{12} is 2. If $\ell=(2,2)$ and $k=1$, then the corresponding element in M_{12} is 0. Define $M^*_{12}=(M_{12}>0)$, i.e., an element in M^*_{12} is equal to 1 if the corresponding element in M_{12} is equal or greater than 1, or 0 otherwise. Then, the equilibrium linear prices $p^*_{\mathbf{J}_1}$ in the counterfactual must satisfy the following FOCs:

$$M^{*\mathrm{T}}_{12}F(M_{12}p^*_{\mathbf{J}_1}) = 0_{|\mathbf{J}_1|\times 1}, \tag{47}$$

where $F(\cdot)$ is defined in (46) and $c_{\mathbf{J}_1\cup\mathbf{J}_2}$ is the vector of marginal costs obtained in Appendix F.2 (we assume that marginal costs are not affected by the pricing scheme). We consider any solution to the nonlinear system of FOCs (47) as the equilibrium counterfactual vector of linear prices $p^*_{\mathbf{J}_1}$. To implement this solution in practice, one can rely on any standard algorithm (e.g., *fsolve* in MATLAB) using as initial guess of $p^*_{\mathbf{J}_1}$ the observed single-unit prices $p_{\mathbf{J}_1}$. Even though possible, in extensive attempts using multiple initial guesses, we never found more than on solution to FOCs (47) in each market.

F.4 Computation of Tables 6, 7, and 8

In this section, we detail the computation of the entries of Tables 6, 7, and 8 which summarize changes due to the counterfactual linear pricing relative to the observed scenario of quantity discounts. Here we only discuss the computation of absolute changes (in \$, liters (L), or grams), but relative changes (in %) are obtained analogously. For results to be more interpretable, we measure absolute changes per household during a year: e.g., change in liters of CSDs purchased in a year by a typical household of size hs. To this purpose, we first predict the absolute changes at the same level of aggregation used in estimation, the shopping trip level by household size, and then multiply these by the average yearly number of shopping trips specific to the household size (single, multi, or average). In the data, the average number of shopping trips in a year is 63.9 for a single-person household, 72.9 for a multi-person household, and 70.7 for an average household.

Price change. We define the changes in prices Δp_{tj} and $\Delta p_{t\mathbf{b}}$ as follows:

$$\Delta p_{tj} = \text{Median}\left\{ \frac{\sum_{j\in\mathbf{J}} p_{tj}^{\text{linear}} - \sum_{j\in\mathbf{J}} p_{tj}^{\text{observed}}}{|\mathbf{J}|}, t \in \mathbf{T}_0 \right\},$$

$$\Delta p_{t\mathbf{b}} = \text{Median}\left\{ \frac{\sum_{\ell\in\mathbf{C}_2} p_{t\ell}^{\text{linear}} - \sum_{\ell\in\mathbf{C}_2} p_{t\ell}^{\text{observed}}}{|\mathbf{C}_2|}, t \in \mathbf{T}_0 \right\},$$

where $\mathbf{T}_0$ is the set of markets in which the three collections of bundles from Table 5 are observed to be purchased by both household sizes, and superscripts "linear" and "observed" refer to the scenarios of counterfactual linear pricing and observed quantity discounts, respectively.

Quantity change. The quantity of CSDs from a collection of bundles $\mathbf{B} \in \{\mathbf{J}, \mathbf{C}_2\}$ by households of size hs in market t is:

$$Q^{hs}(\mathbf{B}) = \sum_{\mathbf{b}\in\mathbf{B}} |\mathbf{b}| \times \mathcal{s}_{t\mathbf{b}}^{hs},$$

where $|\mathbf{b}|$ is the number of units (liters) in bundle $\mathbf{b}$. Then, the quantity change for households of size hs in Table 6 is:

$$\Delta Q^{hs}(\mathbf{B}) = \text{Median}\left\{ \sum_{\mathbf{b}\in\mathbf{B}} |\mathbf{b}| \times \mathcal{s}_{t\mathbf{b}}^{hs,\text{linear}} - \sum_{\mathbf{b}\in\mathbf{B}} |\mathbf{b}| \times \mathcal{s}_{t\mathbf{b}}^{hs,\text{observed}}, t \in \mathbf{T}_0 \right\}.$$

The relative quantity change conditional on purchase for households of size hs is:

$$\Delta Q_{\text{cond}}^{hs} = \text{Median}\left\{ \frac{\sum_{\mathbf{b}\in\mathbf{C}_1} |\mathbf{b}| \times \mathcal{s}_{t\mathbf{b}}^{hs,\text{linear}}}{\sum_{\mathbf{b}\in\mathbf{C}_1} \mathcal{s}_{t\mathbf{b}}^{hs,\text{linear}}} \Bigg/ \frac{\sum_{\mathbf{b}\in\mathbf{C}_1} |\mathbf{b}| \times \mathcal{s}_{t\mathbf{b}}^{hs,\text{observed}}}{\sum_{\mathbf{b}\in\mathbf{C}_1} \mathcal{s}_{t\mathbf{b}}^{hs,\text{observed}}} - 1, t \in \mathbf{T}_0 \right\}$$

and the relative change in the probability of purchase for households of size hs is:

$$\Delta\text{Prob. of Purchase}^{hs} = \text{Median}\left\{ \frac{\sum_{\mathbf{b}\in\mathbf{C}_1} \mathcal{s}_{t\mathbf{b}}^{hs,\text{linear}}}{\sum_{\mathbf{b}\in\mathbf{C}_1} \mathcal{s}_{t\mathbf{b}}^{hs,\text{observed}}} - 1, t \in \mathbf{T}_0 \right\}.$$

To compute the quantity change of (non-)sugary CSDs, denote by $|\mathbf{b}|_s$ the number of sugary CSD units (liters) in bundle $\mathbf{b}$. Then, the quantity change of sugary CSDs in Table 8 is:

$$\Delta Q_{\text{sugary}} = \text{Median}\left\{ \sum_{\mathbf{b}\in\mathbf{C}_1} |\mathbf{b}|_s \times \mathcal{s}_{t\mathbf{b}}^{\text{linear}} - \sum_{\mathbf{b}\in\mathbf{C}_1} |\mathbf{b}|_s \times \mathcal{s}_{t\mathbf{b}}^{\text{observed}}, t \in \mathbf{T}_0 \right\},$$

and that of non-sugary CSDs is:

$$\Delta Q_{\text{non-sugary}} = \text{Median}\left\{\sum_{\mathbf{b}\in\mathbf{C}_1}(|\mathbf{b}|-|\mathbf{b}|_s)\times \mathscr{s}_{t\mathbf{b}}^{\text{linear}} - \sum_{\mathbf{b}\in\mathbf{C}_1}(|\mathbf{b}|-|\mathbf{b}|_s)\times \mathscr{s}_{t\mathbf{b}}^{\text{observed}}, t\in\mathbf{T}_0\right\}.$$

Profit change. The profit change generated by households of size hs is:

$$\Delta\pi^{hs}(\mathbf{B}) = \text{Median}\left\{\sum_{\mathbf{b}\in\mathbf{B}}(p_{t\mathbf{b}}^{\text{linear}} - c_{t\mathbf{b}})\mathscr{s}_{t\mathbf{b}}^{hs,\text{linear}} - \sum_{\mathbf{b}\in\mathbf{B}}(p_{t\mathbf{b}}^{\text{observed}} - c_{t\mathbf{b}})\mathscr{s}_{t\mathbf{b}}^{hs,\text{observed}}, t\in\mathbf{T}_0\right\},$$

where $c_{t\mathbf{b}}$ is the marginal cost of bundle $\mathbf{b}$ in market t (see Appendix F.2).

Compensating variation. In the setting of PONL model (2), income effects enter linearly into the indirect utilities $U_{it\mathbf{b}}$ for all $\mathbf{b}\in\mathbf{C}$. As a consequence, using the definition in (Bhattacharya, 2018, equation 3), for households of size $hs\in\{\text{single}, \text{multi}\}$, the compensating variation is:

$$\text{CV}_t^{hs} = \frac{\text{CS}_t^{hs,\text{observed}} - \text{CS}_t^{hs,\text{linear}}}{\alpha^{hs}}, \tag{48}$$

where $\text{CS}_t^{hs,\text{observed}}$ and $\text{CS}_t^{hs,\text{linear}}$ are defined as:

$$\text{CS}_t^{hs,d} = \ln\left(\sum_{\ell=0}^{J}\left(\sum_{\mathbf{b}'\in\mathbf{N}_\ell}(\omega_{\mathbf{b}'\ell}\exp(\delta_{t\mathbf{b}'}^{hs,d}))^{1/\lambda_\ell^{hs}}\right)^{\lambda_\ell^{hs}}\right) \tag{49}$$

with $d\in\{\text{observed}, \text{linear}\}$. It follows that the average compensating variation across hs's is:

$$\text{CV}_t = \frac{\sum_{hs} w_{hs}\alpha^{hs}\text{CV}_t^{hs}}{\sum_{hs} w_{hs}\alpha^{hs}}. \tag{50}$$

Finally,

$$\text{CV} = \text{Median}\left\{\text{CV}_t, t\in\mathbf{T}_0\right\}.$$

Denote by "single unit" the scenario in which only $(p_{tj}^{\text{observed}})_{j\in\mathbf{J}}$ change to $(p_{tj}^{\text{linear}})_{j\in\mathbf{J}}$ in $(\delta_{tj}^{hs,\text{observed}})_{j\in\mathbf{J}}$, while $(\delta_{t\mathbf{b}}^{hs,\text{observed}})_{\mathbf{b}\in\mathbf{C}_2}$ remain unchanged. Then, for households of size hs, the compensating variation due to the changes in $(p_{tj})_{j\in\mathbf{J}}$ is defined as:

$$\text{CV}_t^{hs,\text{single unit}} = \frac{CS_t^{hs,\text{observed}} - CS_t^{hs,\text{single unit}}}{\alpha^{hs}},$$

while that due to the changes in $(p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$ as $\text{CV}_t^{hs} - \text{CV}_t^{hs,\text{single unit}}$. Their average across household sizes is then defined as in (50).

Compensating Variation/Expenditure. The expenditure on CSDs for households of size hs in market t in the observed scenario of quantity discounts is:

$$\text{Expenditure}_t^{hs} = \sum_{\mathbf{b} \in \mathbf{C}_1} p_{t\mathbf{b}} \mathscr{s}_{t\mathbf{b}}^{hs,\text{observed}}.$$

Then, the median of the ratio CV/Expenditure for households of size hs is:

$$\text{CV/Expenditure}^{hs} = \text{Median}\left\{ \frac{\text{CV}_t^{hs}}{\text{Expenditure}_t^{hs}}, t \in \mathbf{T}_0 \right\},$$

while that across hs's is:

$$\text{CV/Expenditure} = \text{Median}\left\{ \frac{\text{CV}_t}{\sum_{hs} w_{hs} \text{Expenditure}_t^{hs}}, t \in \mathbf{T}_0 \right\}.$$

Predicted added sugar change. Denote by τ_j the added sugar content (grams) in one unit (liter) of CSD j. The added sugar amount (grams) in bundle $\mathbf{b}$ is $\tau_{\mathbf{b}} = \sum_{j \in \mathbf{b}} \tau_j$. Then, the predicted sugar change in Table 8 is:

$$\Delta Q_{\text{added sugar}} = \text{Median}\left\{ \sum_{\mathbf{b} \in \mathbf{C}_1} \tau_{\mathbf{b}} \times \mathscr{s}_{t\mathbf{b}}^{\text{observed}} - \sum_{\mathbf{b} \in \mathbf{C}_1} \tau_{\mathbf{b}} \times \mathscr{s}_{t\mathbf{b}}^{\text{linear}}, t \in \mathbf{T}_0 \right\}.$$

F.5 Understanding Compensating Variations in Table 7

Here we discuss a simple example useful to get some insight about the compensating variations reported in Table 7. Consider a setting with $\mathbf{J} = \{1\}$ and $\mathbf{C}_2 = \{(1,1)\}$. From (49), the consumer surplus for households of size $hs = \{\text{single}, \text{multi}\}$ at prices $(p_1, p_{(1,1)})$ is:

$$\text{CS}^{hs}(p_1, p_{(1,1)}) = \ln\left(1 + \left(\exp\left\{\frac{\delta_1^{hs} - \alpha^{hs} p_1}{\lambda^{hs}}\right\} + \exp\left\{\frac{2\delta_1^{hs} - \alpha^{hs} p_{(1,1)} + \Gamma^{hs}}{\lambda^{hs}}\right\}\right)^{\lambda^{hs}}\right).$$

By a first-order Taylor expansion of (48), the compensating variation associated to a change in prices from $(p_1, p_{(1,1)})$ to $(p_1 + \Delta_1, p_{(1,1)} + \Delta_{(1,1)})$ can be approximated as:

$$\begin{aligned}
\text{CV}^{hs}(\Delta_1, \Delta_{(1,1)}) &= \frac{\text{CS}^{hs}(p_1, p_{(1,1)}) - \text{CS}^{hs}(p_1 + \Delta_1, p_{(1,1)} + \Delta_{(1,1)})}{\alpha^{hs}} \\
&\approx -\frac{1}{\alpha^{hs}}\left[\frac{\partial \text{CS}^{hs}(p_1, p_{(1,1)})}{\partial p_1}\Delta_1 + \frac{\partial \text{CS}^{hs}(p_1, p_{(1,1)})}{\partial p_{(1,1)}}\Delta_{(1,1)}\right] \\
&= \mathscr{s}_1^{hs}\Delta_1 + \mathscr{s}_{(1,1)}^{hs}\Delta_{(1,1)}.
\end{aligned}$$

This shows that the compensating variation due to $\Delta_{(1,1)}$ (or Δ_1) is approximately proportional to hs's probability to purchase $(1,1)$ (or 1) at $(p_1, p_{(1,1)})$. As documented in Figure 3, in the observed scenario

Table 9: Counterfactual Linear Pricing: Percentage Changes in Price, Quantity, and Profit

	Average	Single-person households	Multi-person households
Price Change (%)			
Δp_{tj}	-31.53%		
	(0.79%)		
$\Delta p_{t\mathbf{b}}$	$+14.79\%$		
	(1.68%)		
Quantity Change (%)	-20.38%	-17.64%	-20.99%
	(1.40%)	(1.45%)	(1.45%)
Single units	$+50.20\%$	$+32.45\%$	$+60.65\%$
	(2.17%)	(3.44%)	(3.12%)
Multiple units	-21.06%	-18.70%	-21.83%
	(1.41%)	(1.62%)	(1.51%)
Profit Change (%)	-20.53%	-15.48%	-21.27%
	(3.32%)	(3.82%)	(3.41%)
Single units	$+1.99\%$	-4.04%	$+5.46\%$
	(1.30%)	(1.64%)	(1.39%)
Multiple units	-21.53%	-17.58%	-22.43%
	(3.41%)	(4.20%)	(3.49%)

Notes: We report all the computational details of the above entries in Appendices F.2 (marginal costs), F.3 (counterfactual simulation), and F.4 (price, quantity, and profit changes). All entries are computed as medians over the same set of markets used in Table 5 to compute price elasticities. Standard errors are obtained using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

with quantity discounts, multi-person households are far more likely than single-person households to purchase multiple units of CSDs, so that $\mathscr{s}^{\text{multi}}_{(1,1)} > \mathscr{s}^{\text{single}}_{(1,1)}$ and $\mathscr{s}^{\text{multi}}_{1} < \mathscr{s}^{\text{single}}_{1}$. In addition, our simulated counterfactual suggests that going from quantity discounts to linear pricing would result in $\Delta_1 < 0$ and $\Delta_{(1,1)} > 0$ (Table 6). Combining these observations should clarify the patterns reported in Table 7. In particular, using the simpler notation from the current example: $\text{CV}^{\text{multi}}(\Delta_1, \Delta_{(1,1)}) > \text{CV}^{\text{single}}(\Delta_1, \Delta_{(1,1)})$ because of the larger weight $\mathscr{s}^{\text{single}}_{1}$ single-person households place on $\Delta_1 < 0$ and the larger weight $\mathscr{s}^{\text{multi}}_{(1,1)}$ multi-person households instead place on $\Delta_{(1,1)} > 0$.

F.6 Additional Tables

Table 10: Carbonated Soft Drink Products and their Added Sugar Content

"L5" variable in IRI (Product)	Added Sugar (gr. per L)
7 UP	70
7 UP POMEGRANATE	100
7 UP RETRO	107.08
A & W	121.73
AIRFORCE NUTRISODA	0
BARQS	123.98
CAFFEINE FREE COKE CLASSIC	106
CAFFEINE FREE DIET COKE	0
CAFFEINE FREE DIET DR PEPPER	0
CAFFEINE FREE DIET PEPSI	0
CAFFEINE FREE PEPSI	0
CANADA DRY	98.62

CANFIELD	0
CHERRY 7 UP	107.08
CHERRY COKE	118.35
CHERRY R C	120.83
CHERRY VANILLA DR PEPPER	106.51
CITRUS BLAST	0
COKE CHERRY ZERO	0
COKE CLASSIC	106
COKE ZERO	0
CRUSH	121.17
DIET 7 UP	0
DIET BARQS	0
DIET CHERRY 7 UP	0
DIET CHERRY CHOCOLATE DR PEPP	0
DIET CHERRY COKE	0
DIET CHERRY VANILLA DR PEPPER	0
DIET CITRUS BLAST	0
DIET COKE	0
DIET COKE PLUS	0
DIET COKE WITH LIME	0
DIET COKE WITH SPLENDA	0
DIET DR PEPPER	0
DIET MOUNTAIN DEW	0
DIET MOUNTAIN DEW CAFFEINE FR	0
DIET MOUNTAIN DEW CODE RED	0
DIET MOUNTAIN DEW ULTRA VIOLE	0
DIET MTN DEW SUPERNOVA	0
DIET MTN DEW VOLTAGE	0
DIET MUG	0
DIET PEPSI	0
DIET PEPSI JAZZ	0
DIET PEPSI NFL KICKOFF	0
DIET PEPSI TWIST	0
DIET PEPSI VANILLA	0
DIET PEPSI WITH LIME	0
DIET R C	0
DIET RITE	0
DIET RITE PURE ZERO	0
DIET SCHWEPPES	0
DIET SIERRA MIST	0
DIET SIERRA MIST CRANBERRY SP	0
DIET SIERRA MIST RUBY SPLASH	0
DIET SQUIRT	0
DIET SUN DROP	0
DIET SUNKIST	0
DIET TROPICANA TWISTER	0

DIET VERNORS	0
DIET WILD CHERRY PEPSI	0
DR PEPPER	108.20
DR PEPPER HERITAGE	79.37
DR PEPPER TEN	6.8
FANTA	123.42
FANTA ZERO ORANGE	0
FRESCA	0
GRAFS	135.21
GRAFS 50 50	112.68
HIRES	112.33
IBC	126.12
MELLO YELLO	129.58
MELLO YELLO THE ORIGINAL SMOO	129.58
MELLO YELLO ZERO	0
MOUNTAIN DEW	130.29
MOUNTAIN DEW CAFFEINE FREE	130.18
MOUNTAIN DEW CODE RED	128.49
MOUNTAIN DEW DISTORTION	129.62
MOUNTAIN DEW GAME FUEL	50.71
MOUNTAIN DEW PITCH BLACK	126.80
MOUNTAIN DEW REVOLUTION	126.8
MOUNTAIN DEW SUPERNOVA	129.6
MOUNTAIN DEW THROWBACK	123.5
MOUNTAIN DEW TYPHOON	132.28
MOUNTAIN DEW VOLTAGE	129.62
MOUNTAIN DEW WHITE OUT	128.49
MR PIBB ZERO	0
MT DEW LIVE WIRE	130.18
MUG	120.25
PEPSI	116.28
PEPSI MAX	0
PEPSI MAX CEASE FIRE	0
PEPSI NFL KICKOFF	121
PEPSI ONE	0
PEPSI THROWBACK	113.28
PIBB XTRA	107.08
R C	116.67
RUBY RED SQUIRT	123.98
SCHWEPPES	92.99
SEAGRAMS	97.00
SIERRA MIST	104.91
SIERRA MIST CRANBERRY SPLASH	97.22
SIERRA MIST FREE	0
SIERRA MIST FREE UNDERCOVR OR	0
SIERRA MIST NATURAL	170.83

SIERRA MIST RUBY SPLASH	0
SIERRA MIST UNDERCOVER ORANGE	109.90
SPRITE	107.08
SPRITE ZERO	0
SQUIRT	107.04
SUN DROP	139
SUNKIST	122.4
SUNKIST CITRUS FUSION	132.77
SUNKIST SOLAR FUSION	198.66
TAB	0
TAHITIAN TREAT	138.07
TAVA BRAZILIAN SAMBA	0
TAVA MEDITERRANEAN FIESTA	0
TAVA TAHITIAN TAMURE	0
TROPICANA TWISTER	101.3
VANILLA COKE	118.35
VANILLA COKE ZERO	0
VAULT	129.9
VAULT RED BLITZ	135
VAULT ZERO	0
VERNORS	108.21
WELCHS	138.03
WILD CHERRY PEPSI	123.46
OTHERS	57.94

Notes: For the products still available (10/2021), we collected added sugar information from nutrition labels on producers' websites. For those discontinued, we instead collected added sugar information from nutrition comparison websites. By definition, diet products have zero added sugar. We assume that the residual product "Others" has an amount of sugar per liter equal to the average of the remaining 127 products. All web links to retrieve this information are available on request.