An Empirical Model of Quantity Discounts with Large Choice Sets

Alessandro Iaria and Ao Wang*

November 17, 2022

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 −31.6% while those of larger quantities increase by +14.9%. Purchased quantities decrease by −20.7% and industry profit by −19.74%. Consumer surplus however reduces only moderately, suggesting that a ban on quantity discounts for sugary drinks may be a simple and effective policy to limit added sugar intake. Our calculations confirm that such a ban would indeed be as effective as a sugar tax of 1 cent per ounce of added sugar and reduce added sugar intake by −22.1%.

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

*University of Bristol and CEPR, alessandro.iaria@bristol.ac.uk and University of Warwick, ao.wang@warwick.ac.uk. We thank for the very insightful comments and discussions Simon Anderson, Gaurab Aryal, Steve Berry, Federico Ciliberto, Valentina Corradi, Cristina Gualdani, Phil Haile, Jean-Francois Houde, Joshua Lanier, Jeremy Large, Lorenzo Magnolfi, John Quah, Alan Sorensen, Otto Toivanen, and the participants to seminars and workshops at Asia Meeting of the Econometric Society 2022, Boston University, BSE Summer Forum 2022 (Applied IO), China Meetings of the Econometric Society 2021, Colombia Central Bank, EARIE 2021, IAAE 2022, MaCCI Summer Institute 2021, National University of Singapore, NBER Summer Institute (IO), OCEA Webinar at the Competition and Markets Authority, Queen Mary University of London, Scottish Economic Society 2022, Southern Economic Association 2021, Singapore Management University, University of Glasgow, University of Surrey, University of Virginia, University of Wisconsin-Madison, and Yale University. We would like to thank IRI for making the data used in this paper available. All estimates and analyses in this paper, based on data provided by IRI, are by the authors and not by IRI.

1 Introduction

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

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: the number of ways in which consumers can combine products into bundles grows steeply in the number of products and the number of parameters capturing unobserved synergies among products 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 by proposing a novel method to estimate demand for bundles in the presence of large choice sets, which we then apply to investigate quantity discounts in the market for carbonated soft drinks. We propose a Generalized Nested Logit model, called Product-Overlap Nested Logit (PONL), that has as many overlapping nests as products and where each bundle belongs to all nests corresponding to its product components. Because of the overlapping nests, the PONL model can handle realistic forms of unobserved heterogeneity but cannot however be estimated on the basis of Berry (1994) and, because of the large choice sets, GMM procedures as in Berry et al. (1995) may be impractical. We instead devise an optimization- and derivative-free Gauss-Siedel iterative procedure that can be parallelized over both bundles and markets, virtually eliminating any chal-

[1]In contrast, several US retail chains do not take full advantage of other forms of price discrimination and charge uniform prices for given quantity of a product/service across 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.

lenge of dimensionality due to large choice sets.[3] In an extension, we also show that the PONL model can include a random coefficient on an observable attribute such as price or bundle size, while preserving its practical estimation convenience.

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.[4] 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.[5] We however show that the PONL model can be estimated by augmenting the classic 2SLS linear system with non-linear equations that account for the overlapping nests. Implementation requires that at least some of the products can be purchased in isolation and not only as part of bundles (pure bundling can be handled for some but not all products). The proposed estimator is robust to price endogeneity and is easy to implement with large choice sets. Differently, likelihood-type estimators based on the direct use individual-level purchases would not be convenient with large choice sets, mainly because of the large number of fixed effects required to control for price endogeneity (Grieco et al., 2022; Iaria and Wang, 2019).[6]

As part of a broader anti-obesity strategy, on 6 April 2022 the UK government proposed to ban quantity discounts on unhealthy foods and drinks from supermarkets. The proposed ban was received with strong opposition by packaged food and drinks giants such as Kellogg's, which promptly launched a legal action against the regulation. While on 4 July 2022 Kellogg's court challenge failed, the UK government announced the intention to delay the implementation of the ban to help alleviate the recent cost of living crisis. To inform this ongoing policy debate, we employ our method and investigate the likely welfare consequences of a ban on quantity discounts in the market for carbonated soft drinks (CSDs).

[3]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 expected to improve over time as these become more cheaply available.

[4]By "bundle-level" purchase probability we mean the joint probability that a bundle made of one or more units of the same or of different products is purchased. For example, the probability that two units of product j and one unit of product k are jointly purchased.

[5]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.

[6]In our application, in order to include a fixed effect for each bundle-market combination (as we do with our estimator), one would have to numerically minimize a non-linear likelihood function with more than $176,700$ parameters, which is currently not viable. One could simplify implementation by imposing strong restrictions on these fixed effects, but as we illustrate in our application, sensible restrictions are hard to specify a priori.

Using household-level purchase data by IRI for the period 2008-2011 in the USA, 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 that quantity discounts are pervasive (e.g., the average unit-price of a Diet Coke is higher for a 12oz can than for a 2L bottle). According to intuition, larger households purchase larger quantities of CSDs, both of the same and of different products. Despite being unable to price discriminate directly on the basis of household size, 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). Because also single-person households purchase multiple units of CSDs, quantity discounts however only achieve imperfect screening. In this complex situation of imperfect screening in an oligopolistic market with differentiated products, the welfare effects of a ban on quantity discounts are ambiguous (Anderson and Leruth, 1993; Armstrong, 2013; Varian, 1989).

We estimate a flexible PONL model with around $16,900$ bundles of CSDs and $176,700$ demand parameters and empirically assess the welfare effects of a ban on (observed) quantity discounts by simulating a counterfactual with linear pricing (forcing constant unit-prices for all products). Our counterfactual simulations suggest that linear pricing would lead to a reduction of -31.6% in the average price of small quantities (up to one liter) and to an increase of $+14.9\%$ 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 -20.7% and, as a consequence, industry profit would shrink by -19.7%. Despite the substantial reduction in quantity purchased, consumer surplus would however not reduce too sharply, with a compensating variation of $+3.7$\$ per household-year (amounting to 2.8% 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 +4.1$ as opposed to +1.8$).

These results suggest that a ban on quantity discounts could be a practically simple and effective policy to limit 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 prices 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 calculations confirm that a ban on quantity discounts for sugary CSDs would be as effective as a sugar tax of 1 cent per ounce of added sugar and reduce added sugar intake by −22.1%.[7]

There is a large empirical literature leveraging the estimation of demand for bundles.[8] 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 either focus on applications with restricted choice sets or limited forms of unobserved heterogeneity. Our methods enable empirical researchers to scale up the number of bundles considered while allowing for realistic forms of unobserved heterogeneity, facilitating the investigation of demand across multiple product categories, like grocery or online shopping (Reimers and Waldfogel, 2021; Thomassen et al., 2017), 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), spillovers of taxes from a product

[7] 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).

[8] Some examples are: Crawford and Yurukoglu (2012); Dubé (2004); Florez-Acosta and Herrera-Araujo (2020); Fosgerau et al. (2021); Gentzkow (2007); Hendel (1999); Ho et al. (2012); Manski and Sherman (1980); Ruiz et al. (2020); Thomassen et al. (2017).

category to others (Allcott et al., 2019; Dubois et al., 2020), portfolio choice models of asset pricing (Koijen and Yogo, 2019), and many more.

Three novel approaches to addressing large choice sets in the estimation of demand for bundles were recently proposed by Ershov et al. (2021), Lewbel and Nesheim (2019), and Lanier et al. (2022). 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) depart from discrete choice models and specify a more general discrete-continuous choice model as in Dubin and McFadden (1984). They address large choice sets with sparsity, by assuming that each consumer purchases 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.

Lanier et al. (2022) build on the long-panel framework by Dubois et al. (2020) and propose to estimate individual-specific multinomial logit models of demand for bundles on choice subsets as in McFadden (1978). They allow for more flexible unobserved heterogeneity than the PONL model, but pose more restrictions on the demand synergies among products and do not allow for price endogeneity.

Our empirical analysis contributes to the literature on CSDs. Some of the papers in this literature estimate demands for multiple units (Chan, 2006; Dubé, 2004; Hendel and Nevo, 2013; Wang, 2015), while others focus on different aspects of the industry, such as vertical relations 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 investigate the welfare effects of quantity discounts in this industry.[9]

[9] Bonnet and Dubois (2010) do not study non-linear pricing with respect to "final" consumers (as we do in this paper), but rather two-part tariff contracts between manufacturers and retailers. While close in spirit, Hendel and Nevo (2013) study the welfare effects of intertemporal price discrimination (i.e., temporary price reductions) rather than quantity discounts.

2 The Product-Overlap Nested Logit (PONL)

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; Fox and Lazzati, 2017; Iaria and Wang, 2021; Wang, 2019). Each $\mathbf{b} \in \mathbf{C}_1$ is a combination of 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 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 that 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., 2021), cannot appropriately capture these intuitive patterns of correlation.[10] On the other hand, more appropriate non-parametric (Compiani, 2019) or even mixed logit models can be unfeasible in applications with large choice sets (Gentzkow, 2007; Iaria and Wang, 2019; Liu et al., 2010). As a solution, we propose a special case of Generalized Nested Logit (GNL) model (Abbe et al., 2007; Bierlaire, 2006; Wen and Koppelman, 2001) that specifically accounts for the product overlap in the unobserved preferences of different bundles and is practically convenient with large choice sets. We call it the Product-Overlap Nested Logit (PONL) model. In an extension, we also consider the inclusion of a random coefficient on an observable attribute such as bundle size or price and show that, despite the ad-

[10]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 (only one 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.

ditional complexity, the PONL model is still practically convenient in applications with large choice sets.

2.1 Correlation of Unobserved Preferences 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 a bundle $\mathbf{b}$ that includes at least one unit of j and one of j'. Denote by $U_{it\mathbf{b}}$ the indirect utility of household i from purchasing bundle $\mathbf{b}$ in market t:

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

where $\delta_{t\mathbf{b}}$ is the average utility of $\mathbf{b}$ among the households in market t and $\varepsilon_{it\mathbf{b}}$ is an unobserved component of utility specific to household i. The PONL model is obtained as a special case of the GNL model (Wen and Koppelman, 2001) by assuming that $\varepsilon_{it\mathbf{b}}$ is distributed according to a Generalized Extreme Value (GEV) with generating function (McFadden, 1978):

$$G\left(\exp(\delta_{t\mathbf{b}}), \mathbf{b} \in \mathbf{C}\right) = \sum_{j=1}^{J} \left(\sum_{\mathbf{b}' \in \mathbf{N}_j} \left(\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'})\right)^{1/\lambda_j} \right)^{\lambda_j}, \tag{2}$$

where $\omega_{\mathbf{b}j}$ is a weight, called allocation parameter, that determines to which extent $\mathbf{b}$ belongs to nest $\mathbf{N}_j$ on the basis of its observed product components and $\lambda_j \in (0, 1]$ is a nesting parameter that determines the strength of the correlation among the bundles in $\mathbf{N}_j$.[11] In particular, the PONL model is obtained by defining the allocation parameter $\omega_{\mathbf{b}j}$ as "the proportion of units of product j included in bundle $\mathbf{b}$." Each $\omega_{\mathbf{b}j}$ equals zero if $\mathbf{b} \notin \mathbf{N}_j$ ($\mathbf{b}$ does not include any unit of j) or, if $\mathbf{b} \in \mathbf{N}_j$, it is the proportion of units of j in $\mathbf{b}$. For example, bundle $(1, 1, 2)$ has $\omega_{\mathbf{b}1} = 2/3$, $\omega_{\mathbf{b}2} = 1/3$, and $\omega_{\mathbf{b}j'} = 0$ for any $j' \neq 1, 2$, and is proportionally allocated to nests

[11] The allocation parameters are "weights" in the sense that $0 \leq \omega_{\mathbf{b}j} \leq 1$ for each $(\mathbf{b}, j)$ and $\sum_{j=1}^{J} \omega_{\mathbf{b}j} = 1$ for each $\mathbf{b}$. Bierlaire (2006) shows that any GNL model, and hence the PONL model, is consistent with random utility maximization when $\lambda_j \in (0, 1]$, $j = 1, ..., J$.

$\mathbf{N}_1$ (two out of three units) and $\mathbf{N}_2$ (one out of three units).[12]

The PONL model implies a correlation structure among the unobserved preferences of different bundles that can be intuitively approximated as (Papola, 2004):[13]

$$\mathrm{Corr}(U_{it\mathbf{b}}, U_{it\mathbf{b}'}) \approx \sum_{j=1}^{J} \omega_{\mathbf{b}j}^{1/2} \omega_{\mathbf{b}'j}^{1/2} \left(1 - \lambda_j^2\right), \tag{3}$$

which highlights how the PONL model generalizes the NL model.[14] 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. As 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 nesting parameters proportionally to the overlap in product composition between bundles $\mathbf{b}$ and $\mathbf{b}'$.

Example 1. The possibility of any bundle to belong to multiple nests plays a conceptual role in empirical models of demand for bundles: for each bundle is a combination of products, any product will 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 one unit of 1 and of 2, so that $\mathbf{C}_1 = \{1, 2, 3, (1,1), (2,2), (1,2)\}$.

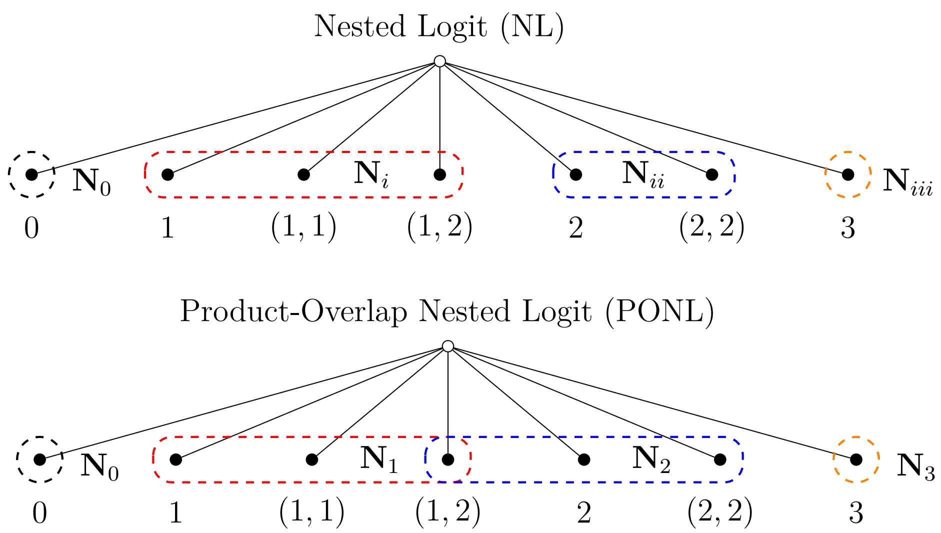
The NL model would require to uniquely and arbitrarily allocate each element of $\mathbf{C}_1$ to a nest. For example, one could specify three nests: $\mathbf{N}_i = \{1, (1,1), (1,2)\}$, $\mathbf{N}_{ii} = \{2, (2,2)\}$, and $\mathbf{N}_{iii} = \{3\}$. While it is true that the alternatives *within* each nest share some common feature, i.e. product 1 in $\mathbf{N}_i$, product 2 in $\mathbf{N}_{ii}$, and

[12] The estimation of GNL models requires the ex-ante specification of the allocation parameters (Bierlaire, 2006), which are too many to be separately identified along the structural parameters. This plays a similar role as the ex-ante specification of the nesting structure in NL models. Our definition of $\omega_{\mathbf{b}j}$ specializes the GNL model to the PONL model and gives rise to the desirable computational advantages described in the paper. For robustness, in the empirical application we experimented with alternative specifications of $\omega_{\mathbf{b}j}$ that preserve the PONL structure, such as $\omega_{\mathbf{b}j} = \mathbf{1}_{\mathbf{b}\in\mathbf{N}_j} \times (\sum_{j'=1}^{J} \mathbf{1}_{\mathbf{b}\in\mathbf{N}_{j'}})^{-1}$, and obtained very similar results.

[13] An exact closed-form expression for $\mathrm{Corr}(U_{it\mathbf{b}}, U_{it\mathbf{b}'})$ has not yet been derived for the GNL model. The approximation in (3) was conjectured by Papola (2004) and, to the best of our knowledge, it is still the most used closed-form approximation put forward in the literature. In simulation studies, Abbe et al. (2007); Marzano and Papola (2008) note that Papola (2004)'s approximation tends to overestimate the true correlation of the GNL model. Marzano et al. (2013) propose a more sophisticated numerical approximation to this correlation, which however is not in closed-form and does not provide intuition on the underlying correlation structure of the PONL model and its relationship with the NL model.

[14] In the NL model, the correlation between any two bundles $\mathbf{b}$ and $\mathbf{b}'$ is $1 - \lambda_j^2$ if both belong to nest j, while it is 0 if the two bundles belong to different nests.

Figure 1: Nesting Structures of NL and PONL



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.[15] 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. 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 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 two 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. ■

[15] 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.

Example 2. To better understand how the allocation parameters proportionally allocate bundles to nests, and the implication of this proportional allocation on the correlation among unobserved preferences, consider an example with four bundles. Bundle $\mathbf{b}$ that only includes units of product j with $\omega_{\mathbf{b}j} = 1$, $\mathbf{b}'$ that only includes units of product j' with $\omega_{\mathbf{b}'j'} = 1$, and two other $\mathbf{b}''$ and $\mathbf{b}'''$ that include different combinations of units of both products j and j' with, respectively, $(\omega_{\mathbf{b}''j} = 0.36,\ \omega_{\mathbf{b}''j'} = 0.64)$ and $(\omega_{\mathbf{b}'''j} = 0.64,\ \omega_{\mathbf{b}'''j'} = 0.36)$. Then, bundle $\mathbf{b}$ will be more correlated with $\mathbf{b}'''$ than with $\mathbf{b}''$, i.e. $0.8(1 - \lambda_j^2) > 0.6(1 - \lambda_j^2)$, given that $\mathbf{b}'''$ includes a larger proportion of units of product j than $\mathbf{b}''$. Symmetrically, bundle $\mathbf{b}'$ will be more correlated with $\mathbf{b}''$ than with $\mathbf{b}'''$, given that $\mathbf{b}''$ includes a larger proportion of units of product j' than $\mathbf{b}'''$. Moreover, because both bundles $\mathbf{b}''$ and $\mathbf{b}'''$ include units of products j and j', $\mathrm{Corr}(U_{it\mathbf{b}''}, U_{it\mathbf{b}'''})$ will depend on both λ_j and $\lambda_{j'}$: $0.48(1 - \lambda_j^2) + 0.48(1 - \lambda_{j'}^2)$. 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. ■

2.1.1 Extension: A Random Coefficient on an Observable Attribute

The PONL model can be extended to allow for additional sources of correlation in unobserved preferences on the basis of observable attributes other than product components. For example, households may have idiosyncratic preferences for observable attributes such as bundle size (e.g., larger households may need to purchase larger quantities) or the price of a bundle (e.g., heterogeneous marginal utility of income). We consider an extension of (1) that includes a random coefficient ν_i on an observable attribute such as the price of a bundle, $p_{t\mathbf{b}}$:

$$\begin{aligned} U_{it\mathbf{b}}(\sigma) &= \delta_{it\mathbf{b}} + \varepsilon_{it\mathbf{b}} \\ &= \delta_{t\mathbf{b}} - \sigma \nu_i p_{t\mathbf{b}} + \varepsilon_{it\mathbf{b}}, \end{aligned} \tag{4}$$

so that the covariance among the unobserved preferences of any two bundles $\mathbf{b}$ and $\mathbf{b}'$,[16] $\mathrm{Cov}(U_{it\mathbf{b}}(\sigma), U_{it\mathbf{b}'}(\sigma)) = \sigma^2\, p_{t\mathbf{b}}\, p_{t\mathbf{b}'} + \mathrm{Cov}(U_{it\mathbf{b}}(0), U_{it\mathbf{b}'}(0))$ with $U_{it\mathbf{b}}(0)$ denoting $U_{it\mathbf{b}}(\sigma)$ evaluated at $\sigma = 0$, also depends on prices, through $\sigma^2\, p_{t\mathbf{b}}\, p_{t\mathbf{b}'}$, in addition to the overlap in product components, through $\mathrm{Cov}(U_{it\mathbf{b}}(0), U_{it\mathbf{b}'}(0))$ (which does not depend on prices). Unobserved heterogeneity in price sensitivity, i.e. $\sigma > 0$, implies $\mathrm{Cov}(U_{it\mathbf{b}}(\sigma), U_{it\mathbf{b}'}(\sigma)) > \mathrm{Cov}(U_{it\mathbf{b}}(0), U_{it\mathbf{b}'}(0))$ due to $\sigma^2\, p_{t\mathbf{b}}\, p_{t\mathbf{b}'} > 0$, while larger values of σ lead to a more prominent role played by prices relative to product overlap in determining $\mathrm{Cov}(U_{it\mathbf{b}}(\sigma), U_{it\mathbf{b}'}(\sigma))$. In the interest of space, in

[16] Assuming that ν_i is independent of the unobserved GEV term $\varepsilon_{it\mathbf{b}}$.

what follows we focus on the identification and estimation of PONL model (1) and its novel features, while in Appendix E we sketch the extensions of our arguments to the inclusion of a random coefficient ν_i as in (4).

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 assume 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.[17] 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:

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

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

[17] In Appendix G, 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 5.

et al., 2017) or aggregation across multiple choices (Dubé, 2004; Hendel, 1999).[18]

In the context of quantity discounts, for example, even excluding any other source of synergy, $\Gamma_{t\mathbf{b}} = -\alpha(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}) > 0$ whenever it is cheaper to purchase the products in bundle $\mathbf{b}$ jointly rather than separately and $p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj} < 0$. In this case, a random coefficient on price as in (4) would introduce household-level unobserved heterogeneity in the demand synergies: $\delta_{it\mathbf{b}} = \sum_{j\in\mathbf{b}} \delta_{itj} + \Gamma_{it\mathbf{b}} = \sum_{j\in\mathbf{b}} \delta_{tj} - \sigma\nu_i \sum_{j\in\mathbf{b}} p_{tj} + \Gamma_{it\mathbf{b}}$, with $\Gamma_{it\mathbf{b}} = -(\alpha+\sigma\nu_i)(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj})$, allowing for the possibility that more price sensitive households find quantity discounts relatively more attractive, i.e. $\Gamma_{it\mathbf{b}} > \Gamma_{t\mathbf{b}}$.

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 $s_{t\mathbf{b}}$, $s_{t(\mathbf{b}|j)}$, s_t^j, and s_{t0} the t-specific purchase probabilities of, respectively: $\mathbf{b}$, $\mathbf{b}$ 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 $s_{t\mathbf{b}} = s_{t(\mathbf{b}|j)} s_t^j$. Any $\mathbf{b}$ that instead belongs to multiple nests has $s_{t\mathbf{b}}$ given by the sum of the joint purchase probabilities $s_{t(\mathbf{b}|j)} s_t^j$ over the $J+1$ nests, where $s_{t(\mathbf{b}|k)} = 0$ for any k such that $\mathbf{b} \notin \mathbf{N}_k$.

Given (1) and generating function (2), the PONL purchase probability of $\mathbf{b} \in \mathbf{C}_1$ in market t is (McFadden, 1978; Wen and Koppelman, 2001):

$$s_{t\mathbf{b}} = \sum_{j=0}^{J} \underbrace{\frac{(\omega_{\mathbf{b}j} \exp(\delta_{t\mathbf{b}}))^{1/\lambda_j}}{\sum_{\mathbf{b}'\in\mathbf{N}_j} (\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j}}}_{s_{t(\mathbf{b}|j)}} \underbrace{\frac{\left(\sum_{\mathbf{b}'\in\mathbf{N}_j} (\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_j}\right)^{\lambda_j}}{\sum_{\ell=0}^{J} \left(\sum_{\mathbf{b}'\in\mathbf{N}_\ell} (\omega_{\mathbf{b}'\ell} \exp(\delta_{t\mathbf{b}'}))^{1/\lambda_\ell}\right)^{\lambda_\ell}}}_{s_t^j}. \tag{7}$$

Because the outside option belongs to its own singleton nest $\mathbf{N}_0$, by further assum-

[18] 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, the need to accommodate different requests may lead to the purchase of multiple units of the same or of different products on any shopping trip.

ing $\delta_{t0} = 0$ we obtain:[19]

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

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 of any bundle $\mathbf{b} \in \mathbf{C}_2$:

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

Different from a NL model, however, the possibility of overlapping nests leads λ_j, $j \in \mathbf{b}$, to be non-linear in (9). In the GNL model, analogous non-linearities 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:

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

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

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

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

3 Identification

A distinctive feature of our approach is the use of different parts of system (9) to identify different parameters 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 of $(\delta, \beta, \alpha, \lambda)$, and then rely on the remaining $T \times C_2$ non-linear

[19] 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.

equations in (12), corresponding to the purchases of *multiple units*, for the identification of the demand synergies $(\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$, $t \in \mathbf{T}$. Alternatively, one could identify the entire PONL model simultaneously from the $T \times C_1$ non-linear equations in (9). While both approaches are possible, we pursue the former because the latter leads to a more complex problem of endogeneity (see next section) and practically less convenient estimators, especially with large choice sets (see 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 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 $(p_{tj}, s_{t(j|j)})$ with ξ_{tj}, and can be addressed by instruments that satisfy moment conditions at the level of the single unit j.[20] 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, suppose to observe prices and characteristics both of single units $(p_{tj}, x_{tj})_{j=1}^{J}$ and of bundles $(p_{t\mathbf{b}}, x_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$.[21] Then the t-specific average utility of $\mathbf{b}$ 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:

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

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

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

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 parameters, moment conditions (6) would not be sufficient for the bundle-level characteristics $(x_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$

[20] 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 $\ln(s_{tj}/s_{t0})$ on $s_{t(j|j)}$ is typical of NL models (independently of price endogeneity), see Berry (1994).

[21] 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.

to be exogeneous in each market t. 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 (6) plus availability of $J+1$ valid instruments will suffice to address the endogeneity of p_{tj} and $\mathcal{s}_{t(j|j)}$ 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 $(\mathcal{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$ and characteristics of single units $(x_{tj}, p_{tj})_{j=1}^{J}$ across T markets, with $T \to \infty$.[22] Note that, if the within-nest purchase probabilities $(\mathcal{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$ 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 $\mathcal{s}_{t(j|j)}$) and then Γ_t from non-linear system (12) by a plug-in. However, the overlapping nesting structure of the PONL model prevents the observability of the within-nest purchase probabilities $(\mathcal{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$, which in turn requires a different identification and estimation strategy.

Example 3. 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 $(\mathcal{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$, one could directly obtain each within-nest purchase probability as $\mathcal{s}_{t(\mathbf{b}|g)} = \mathcal{s}_{t\mathbf{b}} / (\sum_{\mathbf{b}'\in\mathbf{N}_g} \mathcal{s}_{t\mathbf{b}'})$, $g = i, ii, iii$. Differently, because in the PONL model some $\mathbf{b}$ belongs to multiple nests, one cannot determine the

[22] 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.

within-nest purchase probabilities from the observed $(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1}$.[23] 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:

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

preventing 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 (see next section) and how to obtain a computationally convenient estimator (see section 4). We start by defining

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

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:

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

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

[23] 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 $\mathbf{C}_1$.

$(1-\lambda_j)\ln\pi_{tj}$ in the definition of π_{tj} and obtain:

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

Given λ and $\jmath_t$, (16) and (17) define a system of C_1 equations in C_1 unknowns (Γ_t and π_t) for each t. Because each within-nest purchase probability $\jmath_{t(\mathbf{b}|j)}$ is a function of Γ_t and π_t (see the last equality in (7)), one can address the lack of observability of $(\jmath_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1,j\in\mathbf{J}}$ by expressing Γ_t and π_t in terms of λ and $\jmath_t$. To summarize, PONL model (11) and (12) implies:

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

While the presence of (16) and (17) complicates estimation, it basically does not affect identification. The first equation in (18), which is the one we rely on for the identification of of $(\delta,\beta,\alpha,\lambda)$, is subject to the same endogeneity concerns as (11): both p_{tj} and $\frac{\jmath_{tj}/\jmath_{t0}}{\pi_{tj}}$ are functions of the unobserved residuals $\xi_t=(\xi_{tj})_{j=1}^J$. Suppose that a vector of Q instruments z_{tj} with $Q\geq J+1$ were available, and that they satisfied 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).[24]

Proposition 1 (Identification)**.** *Suppose that Assumption 1 in Appendix A.1 holds. 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 $T\to\infty$.

[24]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).

One can prevent incidental parameters problems by relying on (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 non-linear 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. The concentration of (Γ_t, π_t) for each t motivates the name of the estimator we propose in section 4, the *Concentrated* 2SLS.

In Appendix E, we discuss how Proposition 1 can be extended to the inclusion of a random coefficient on price as in (4). The main complication of this extension is the presence of an additional non-linear function of σ in the right-hand side of the first equation in (18).[25] However, σ can be treated similarly to λ and, following Berry and Haile (2014), identification can be obtained given the availability of valid instruments for p_{tj}, $\mathscr{s}_{t(j|j)}$, and the additional non-linear function of σ.

3.3 Choice of Instruments

In this section, we provide some practical guidance on the selection of valid instruments. Different categories of instruments were proposed in the literature to address, respectively, the endogeneity of p_{tj} and that of $\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}}$ 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 (Hausman, 1996; Nevo, 2001). 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 (Berry et al., 1995).

Despite the lack of observability of π_{tj}, appropriate instruments for $\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}}$ can be selected on the basis of their correlation with $\mathscr{s}_{t(j|j)}$. To see this, denote by $\pi_t = (\pi_j(\lambda; \mathscr{s}_t))_{j \in \mathbf{J}}$ a solution to (16) and (17) for given λ and $\mathscr{s}_t$.[26] By a first-order Taylor approximation of $\ln(\pi_{tj}) = \ln(\pi_j(\lambda; \mathscr{s}_t))$ in (18) around its true value $\ln(\pi^0_{tj}) = \ln(\pi_j(\lambda^0; \mathscr{s}_t))$, we obtain:

$$\ln \mathscr{s}_{tj} - \ln \mathscr{s}_{t0} \quad = \delta_j + x_{tj}\beta - \alpha p_{tj} + (1 - \lambda_j) \left[\ln \mathscr{s}_{t(j|j)} - \frac{1}{\mathscr{s}^j_t} \frac{\partial \pi_j(\lambda^0; \mathscr{s}_t)}{\partial \lambda} (\lambda - \lambda^0) \right] + \xi_{tj}, \quad (20)$$

[25]Namely, the Taylor expansion of $\ln(\mathscr{s}_{tj}/\mathscr{s}_{t0})$ with respect to σ around $\sigma = 0$.

[26]By substituting (16) into (17), for given λ and $\mathscr{s}_t$, π_t is the only argument of the resulting system. As a consequence, we can express π_t in terms of λ and $\mathscr{s}_t$. See Assumption 2 in Appendix B for more details.

where the leading term of the first-order Taylor expansion is $\ln \mathcal{s}_{t(j|j)}$.[27] Similar to a scenario in which π_{tj} were observed, a valid instrument is then "something" that shifts $\mathcal{s}_{t(j|j)}$ independently of ξ_{tj} (Berry, 1994). From (7), we also note that:

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

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, we can see that appropriate instruments for $\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_{tj}}$ may be obtained 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 $\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_{tj}}$ as long as nests j and k are overlapping, $\mathbf{N}_j \cap \mathbf{N}_k \neq \emptyset$ (there are bundles including both units of j and of 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 $\mathcal{s}_{t(j|j)}$ through $\delta_{t\mathbf{b}'} - \delta_{tj}$ (Gandhi and Houde, 2019). 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, moment conditions (6) are sufficient also for the validity of these 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}.[28] In Appendix E, we illustrate that similar ideas hold in a PONL model that includes a random coefficient on price and discuss how to select valid instruments for the identification of both λ_j and σ.

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).[29] This could be obtained on the basis of purchase probabilities (7)-(8) and moment conditions (14), and relying either on the fixed point approach (Berry et al., 1995) or on the MPEC approach

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

[28] 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).

[29] 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).

(Dubé et al., 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. The fixed point implementation would require the computation of T demand inverses with C_1 equations each (and no available contraction mapping results), while the MPEC implementation would require the computation of up to $T \times C_1$ non-linear constraints and their derivatives.

To overcome this challenge, we follow our identification strategy and propose a Concentrated 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). The C2SLS can be implemented by a convenient Gauss-Siedel 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.

4.1 A Concentrated Two Stage Least Square (C2SLS)

Our identification strategy leads to a simple estimator based on (18) and moment conditions (19). Define $Z = (Z_t)_{t=1}^T$, $X = \left((e_j)_{j=1}^J, x_t, -p_t, \left(\ln\left(\frac{s_{tj}/s_{t0}}{\pi_{tj}}\right)\right)_{j=1}^J\right)_{t=1}^T$, where e_j is a vector of zeros with a 1 in the j^{th} position, and $Y = \left((\ln s_{tj} - \ln s_{t0})_{j=1}^J\right)_{t=1}^T$. Given $(\pi_t)_{t=1}^T$, we rely on the first equation of (18) to construct the finite-sample counterpart of the moment conditions implied by (19):

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

and obtain estimates of $(\delta, \beta, \alpha, 1-\lambda)$ by minimizing $\|m(\delta, \beta, \alpha, \lambda)\|^2$. From the first-order conditions of this minimization, we obtain

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

which is the 2SLS estimator of $(\delta, \beta, \alpha, 1-\lambda)$ on the basis of instruments Z.[30] Because $(\pi_t)_{t=1}^T$ in X is unknown, we augment this 2SLS by the second and the third equations in (18) and obtain the proposed estimator, the *Concentrated* 2SLS

[30] One could also use criterion functions other than the square of the Euclidean norm of $m(\cdot)$ and obtain different estimators of $(\delta, \beta, \alpha, 1-\lambda)$. We rely on this specific estimator because of its practical convenience in implementation.

(C2SLS), as a solution to the following non-linear system:[31]

$$\begin{cases} (\delta, \beta, \alpha, 1-\lambda)^{\mathrm{T}} = \left(X^{\mathrm{T}}(ZZ^{\mathrm{T}})X\right)^{-1}\left(X^{\mathrm{T}}(ZZ^{\mathrm{T}})Y\right) \\ \text{non-linear equations (16): } \Gamma_t = \left(\Gamma_{\mathbf{b}}(\Gamma_{t\mathbf{b}}; \lambda, \pi_t, \mathcal{s}_t)\right)_{\mathbf{b}\in\mathbf{C}_2} \\ \text{non-linear equations (17): } \pi_t = \left(\phi_j(\pi_t; \lambda, \Gamma_t, \mathcal{s}_t)\right)_{j\in\mathbf{J}}. \end{cases} \tag{22}$$

We 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). To see this, suppose that the within-nest purchase probabilities $(\mathcal{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$ 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 non-linear system (12) by a plug-in. Differently, with overlapping nests $(\mathcal{s}_{t(\mathbf{b}|j)})_{\mathbf{b}\in\mathbf{C}_1, j\in\mathbf{J}}$ are unobserved but must satisfy both the *linear* and the *non-linear* equations in (22).

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 non-linear system (22). With large C, the non-linear 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 challenge by proposing an iterative procedure that does not directly solve non-linear system (22), but only executes a sequence of 2SLS estimators and parallelizable plug-in operations. Together, these simple

[31] The "concentrated" in the name of the estimator refers to concentrating $(\Gamma_t, \pi_t)_{t=1}^T$ out on the basis of (16) and (17) for any given $(\delta, \beta, \alpha, 1-\lambda)$, along the lines of the classic concentrated MLE routinely used in the estimation of panel data models with fixed effects.

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:[32]

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 non-linear 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).[33]

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

Step 2 of the algorithm leverages on the observation that, for any given π_t, the linear equations in (22) are the 2SLS by Berry (1994). Then, steps 1 and 3 update the values of $(\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$ and π_t instead of solving non-linear equations (16) and (17). Step 2 only requires the estimation of a 2SLS, while steps 1 and 3 a parallelizable sequence of plug-ins (not involving numerical optimizations or derivatives).[34]

The proposed algorithm mimics the classic Gauss-Seidel method for solving systems of equations (Hallett, 1982). While similar algorithms were shown to practically facilitate the implementation of linear (Guimaraes and Portugal, 2010) and non-linear fixed effects estimators (Hospido, 2012; Mugnier and Wang, 2022), little is known about their numerical convergence.[35] We next establish that whenever our sequence of regressions and plug-ins numerically converges, it will attain the C2SLS estimator.

[32]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.

[33]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 $(t, \mathbf{b})$.

[34]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.

[35]One of the very few exceptions is Mugnier and Wang (2022), who provide numerical convergence results for two-way fixed effects non-linear panel models.

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,* $\left(\delta^{(\bar{K})}, \beta^{(\bar{K})}, \alpha^{(\bar{K})}, \lambda^{(\bar{K})}\right)$ *and* $\left(\pi_t^{(\bar{K})}, \Gamma_t^{(\bar{K})}\right)_{t=1}^{T}$ *converge to the C2SLS.*

Proof. See Appendix D. □

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.[36] 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).

In Appendix F, we investigate by simulation the finite sample performance of the C2SLS estimator as a function of the choice set size C and the number of iterations $\bar{K}$ in the proposed algorithm. Our results highlight that as few as five iterations can be sufficient for the proposed algorithm to numerically converge and deliver precise estimates. Importantly, the convergence of the algorithm holds irrespective of C, confirming that a few iterations may be enough to implement the C2SLS estimator also in empirical applications with large choice sets.

4.3 Discussion

An essential feature 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,

[36] 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 Appendix F, our Monte Carlo simulations suggest that 5 iterations can already be sufficient to achieve this form of numerical convergence.

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 here, 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 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; Ruiz et al., 2020). However, 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 (Grieco et al., 2022; Iaria and Wang, 2019). The inclusion of a fixed effect for each bundle-market combination (as in the C2SLS estimator) would require to numerically minimize a non-linear likelihood function with more than $T \times C_1$ parameters. One could of course impose strong restrictions on the demand synergies to drastically reduce their number, but sensible restrictions on these parameters are hard to specify a priori.

For example, in our application around 90% of purchases involve multiple units of the same product j, so that around 90% of the demand synergy parameters are of the type $\Gamma_{t\mathbf{b}} = \Gamma_{t(j,\ldots,j)}$. One may then suppose that the indirect utility of multiple units of j is simply linear in the number units, $\delta_{t(j,\ldots,j)} = |(j,\ldots,j)| \times \delta_{tj}$, so that $\Gamma_{t(j,\ldots,j)} = 0$ (no demand synergies, linear marginal utility in quantity), or concave in the number of units, $\delta_{t(j,\ldots,j)} = \delta_{tj} \times |(j,\ldots,j)|^{\rho}$ with $\rho \in (0,1)$, so that $\Gamma_{t(j,\ldots,j)} < 0$ (negative demand synergies, decreasing marginal utility in quantity).[37] This would drastically reduce the number of fixed effects to $T \times J$ (only product-market specific fixed effects), and in some cases (with small T and/or J) allow estimation by maximum likelihood. However, as we illustrate below, the estimated $\Gamma_{t(j,\ldots,j)}$ are largely different from zero and heterogeneous across t and j, following patterns hard to encapsulate in sensible a priori restrictions.

In Appendix E, we discuss how to extend the above C2SLS estimator and iterative procedure to cases in which the PONL model includes a random coefficient ν_i as in (4). The presence of ν_i introduces a non-linearity with respect to σ on the right-hand side of the first equation in (18). As a result, the update of $\left(\delta^{(k)}, \beta^{(k)}, \alpha^{(k)}, \lambda^{(k)}\right)$ and of $\sigma^{(k)}$ in step 2 of the iterative procedure must rely

[37] This is the route followed by models of demand for multiple units as Dubé (2004); Hendel (1999), where consumers can purchase multiple units of the same product with a utility function concave in the number of units that implies negative demand synergies decreasing in quantity.

on GMM rather than on 2SLS. Importantly, despite this additional non-linearity, the implementation of steps 1 and 3 will retain its independence across $(t, \mathbf{b})$ and overall numerical convenience.

5 Quantity Discounts and CSDs

We implement our method to investigate quantity discounts in the market for carbonated soft drinks (CSDs) in the USA. Relying on household-level purchase data from the period 2008-2011 and the PONL model, we assess the welfare effects of a ban on quantity discounts by simulating a counterfactual with linear pricing.

5.1 Data, Definitions, and Descriptive Statistics

We use household-level and store-level IRI data on CSDs for the cities of Pittsfield and Eau Claire (USA) in the period 2008-2011. We refer the reader to Bronnenberg et al. (2008) for a detailed description of these data. We focus on the $I = 6,155$ households observed to purchase CSDs at least once between 2008 and 2011. For these, we observe household size and $1,736,012$ household-level shopping trips to 22 different grocery stores over 208 weeks. A shopping trip is a household's purchase occasion to a grocery store in a given day: during 23.71% of these, CSDs are observed to be purchased. We consider a household to choose the outside option whenever no CSD is purchased during a shopping trip.

We observe households to purchase CSDs on average every 2.22 weeks, suggesting that, on average, they deplete their stocks of CSDs in approximately two weeks. We then define a market $t = 1, ..., 1197$ as a (four weeks $\times$ store) combination to ensure that observed purchases correspond to consumption within the same interval of time. This mitigates concerns about stockpiling, where households buy more "now" for "later" (Hendel and Nevo, 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. The top two panels of Table 1 summarize this information.

We discretize quantities in units of one liter (L): purchases 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 observed in the data. We denote a bundle $\mathbf{b}$ as any combination of units of the same and of different CSDs

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%

we observe to be purchased during any one shopping trip.[38] We refer to "units" or "liters" interchangeably and call "bundles" also the purchases of single units. On average, we observe households to purchase 117.24 units of CSDs per year.[39]

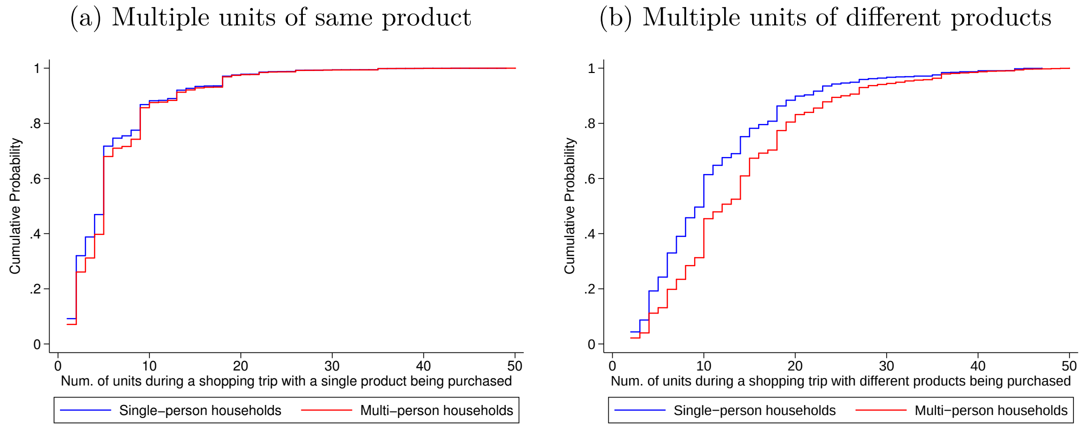
As summarized in the bottom panel of Table 1, we observe $16,873$ different bundles to be purchased during any shopping trip in any market, with an average of 123.8 different bundles within a market. As is well known (Chan, 2006; Dubé, 2004; Ershov et al., 2021), the purchase of multiple units (6.99 on average) is common 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. We divide households into two groups: single-person or

[38]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 G, we however discuss a simple procedure to extend the use of the C2SLS estimator to cases like this where some product is only observed to be purchased through bundles.

[39]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.

multi-person, $hs \in \{\text{single, multi}\}$. Figure 2 shows that multi-person households tend to purchase bundles of larger sizes than single-person households. Because of this, we allow for the possibility that households of different sizes react differently to quantity discounts: we compute purchase probabilities $s_{t\mathbf{b}}^{hs}$ conditional on hs and allow different household sizes to have different demand parameters.[40]

Figure 2: Number of Purchased Units, Single- and Multi-Person Households



We compute each bundle-level price $p_{t\mathbf{b}}$ as the average observed price (in \$) across all shopping trips in t corresponding to purchases of $\mathbf{b}$. Note that third-degree price discrimination cannot be implemented in this context and, within each market, households of different sizes face the same prices. Because IRI records the average store-week price of each UPC separately, we do not observe non-linear prices across UPCs of different products (e.g., joint purchase of 2L Coke Classic and 2L Sprite) and 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 on quantity discounts, both at the disaggregate UPC level (first two columns) and at the product level (last two columns). We regress the unit-price (price per liter) of each UPC/product in a specific store-week on the volume in liters of the UPC/product and fixed effects.[41] Purchases of larger quantities of CSDs are associated to lower prices per unit. For example, a 2L UPC has, on average, a price per liter of -0.267\$ lower than a 1L UPC. Different producers appear to offer comparable quantity discounts.

[40] We compute each $s_{t\mathbf{b}}^{hs}$ as the proportion of shopping trips in t corresponding to purchases of $\mathbf{b}$ by households of size hs.

[41] Each UPC or product corresponds to a specific volume of carbonated soft drink, e.g. 0.33L or 1.5L. We then compute the unit-price of each UPC by dividing its price by its specific volume. For the the product level regressions, we average the unit-prices of the UPCs within each product.

Table 2: Descriptive Evidence on Quantity Discounts

Price per unit ($ per liter)	UPC level		Product level	
Volume in Liters				
Overall	-0.267 (0.001)		-0.235 (0.002)	
Coca-Cola		-0.246 (0.001)		-0.205 (0.002)
PepsiCo		-0.302 (0.001)		-0.273 (0.002)
Dr. Pepper		-0.277 (0.001)		-0.231 (0.002)
Others		-0.245 (0.001)		-0.243 (0.003)
Product fixed effects	yes	yes	yes	yes
Store fixed effects	yes	yes	yes	yes
Time fixed effects	yes	yes	yes	yes
Package type fixed effects	yes	yes	yes	yes
Num. of Obs.	872,532	872,532	206,728	206,728
R^2	0.650	0.652	0.880	0.880

Notes: The first two columns report evidence of quantity discounts at the UPC and store-week level for all UPCs from the store-level IRI data on CSDs for the cities of Pittsfield and Eau Claire in the period 2008-2011. The last two columns repeat the analysis at the product level, defined according to the "brand" variable L5 in the IRI data. All regressions control for package type indicators (plastic or glass bottle, aluminium can, etc.). In the last two columns, these indicators are averaged among the UPCs belonging to each product within a specific store-week.

5.2 Model Specification

In this section, we specify our PONL model of demand for multiple units of CSDs. The average utility of a household of size $hs \in \{\text{single, multi}\}$ 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 researcher. The household size-specific nesting parameter for nest j is $\lambda_j^{hs} = \lambda_{\text{Producer}(j)}^{hs}$, where $\text{Producer}(j) \in \{\text{Coca-Cola/PepsiCo}, \text{Dr.Pepper/Others}\}$ depending on product j's producer.

We use Hausman-type instruments (Hausman, 1996; Nevo, 2001) for the endogenous price p_{tj} and within-nest market share $\mathscr{s}_{t(j|j)}^{hs}$. 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 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. 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.[42]

As mentioned in section 2.2, the demand synergy parameter $\Gamma^{hs}_{t\mathbf{b}}$ captures—among other things—any indirect utility deviation due to non-linear 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:

$$\Gamma^{hs}_{t\mathbf{b}} = -\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right) + \gamma^{hs}_{t\mathbf{b}}, \tag{24}$$

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). As we illustrate next, interpreting the estimated demand synergies requires some care due to the 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. The normalization $\delta^{hs}_{t0} = 0$ consists in subtracting δ^{hs}_{t0} from each $\delta^{hs}_{t\mathbf{b}}$, $\mathbf{b} \in \mathbf{C}_1$. As a result, the identified indirect utility of household size hs from purchasing $\mathbf{b}$ corresponds to $\tilde{\delta}^{hs}_{t\mathbf{b}} = \delta^{hs}_{t\mathbf{b}} - \delta^{hs}_{t0}$ and, in turn, the identified demand synergies to (where $|\mathbf{b}|$ denotes the number of units in $\mathbf{b}$):

$$\begin{aligned}\tilde{\Gamma}^{hs}_{t\mathbf{b}} = \tilde{\delta}^{hs}_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} \tilde{\delta}^{hs}_{tj} &= -\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right) + \gamma^{hs}_{t\mathbf{b}} + (|\mathbf{b}|-1)\delta^{hs}_{t0} \\ &= -\alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right) + \tilde{\gamma}^{hs}_{t\mathbf{b}}.\end{aligned} \tag{25}$$

One can then identify $\tilde{\gamma}^{hs}_{t\mathbf{b}}$, but cannot separately identify $\gamma^{hs}_{t\mathbf{b}}$ and δ^{hs}_{t0} without further assumptions. While this may complicate the interpretation of the estimated demand synergies, in that they will be "shifted" by $(|\mathbf{b}|-1)\delta^{hs}_{t0}$, all the objects of interest (e.g., demand elasticities, marginal costs, consumer surplus, etc.) necessary to perform our counterfactual simulations are only functions of $\tilde{\gamma}^{hs}_{t\mathbf{b}}$—rather than of its individual components—and thus identified.

[42]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.

Table 3: Demand Estimates, Price Coefficients and Nesting Parameters

	(i): $\lambda_j^{hs} = \lambda$	(ii): $\lambda_j^{hs} = \lambda^{hs}$	(iii): $\lambda_j^{hs} = \lambda_{\text{Producer}(j)}^{hs}$
Price coefficients			
α^{single}	0.7455 (0.1061)	0.7415 (0.1061)	0.7448 (0.1073)
α^{multi}	1.0096 (0.0972)	1.0134 (0.0975)	1.0262 (0.0990)
Nesting parameters			
λ	0.8849 (0.0242)		
λ^{single}		0.8726 (0.0283)	
λ^{multi}		0.8900 (0.0255)	
$\lambda^{\text{single}}_{\text{Coca-Cola/PepsiCo}}$			0.8566 (0.0301)
$\lambda^{\text{single}}_{\text{Dr.Pepper/Others}}$			0.9027 (0.0453)
$\lambda^{\text{multi}}_{\text{Coca-Cola/PepsiCo}}$			0.9045 (0.0270)
$\lambda^{\text{multi}}_{\text{Dr.Pepper/Others}}$			0.8527 (0.0379)
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 λ_j^{hs} from the iterative procedure described in section 4. Standard errors are computed using the asymptotic formula detailed in Appendix B.1.

5.3 Estimation Results

Table 3 reports the C2SLS estimates of (23) and λ_j^{hs} 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.[43] The three columns of Table 3 summarize estimation results for three specifications of λ_j^{hs}. 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 $\lambda_j^{hs} = \lambda^{hs}$, while in column (iii) for four $\lambda_j^{hs} = \lambda_{\text{Producer}(j)}^{hs}$. Standard errors are computed using the asymptotic formula 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 the 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., $(1-\lambda_j^{hs}) \approx 0.1/0.15$).[44] Following the procedure in Appendix E, we also estimate a specification as in column (iii) but with a random coefficient on price $\alpha_i^{hs} = \alpha^{hs} + \sigma^{hs}\nu_i$, where ν_i is

[43] Using the criterion from footnote 36, we achieve numerical convergence in 25 iterations.

[44] Despite the unconstrained estimation, all nesting parameters lie between 0 and 1, as required by consistency with utility maximizing behavior (Bierlaire, 2006).

distributed standard normal and σ^{hs} captures the standard deviation of α_i^{hs} among households of size hs. We obtain similar parameter estimates as in column (iii) and variances $(\hat{\sigma}^{\text{single}})^2 = 0.05$ and $(\hat{\sigma}^{\text{multi}})^2 = 0.08$, which do not lead to qualitative changes in our counterfactual simulations. We then rely on the estimates from column (iii), Table 3, as our preferred specification.[45]

Figure 3 plots the estimated demand synergies net of quantity discounts $\tilde{\gamma}_{t\mathbf{b}}^{hs} = \tilde{\Gamma}_{t\mathbf{b}}^{hs} + \alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$, 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. All panels show (left y-axis) the distribution of the estimated $\tilde{\gamma}_{t\mathbf{b}}^{hs}$ for bundles of two, six, and twelve units and (right y-axis) the average of the estimated $\tilde{\gamma}_{t\mathbf{b}}^{hs}$ as a function of $|\mathbf{b}|$, the number of units in $\mathbf{b}$. Despite our inability to directly interpret $\tilde{\gamma}_{t\mathbf{b}}^{hs}$ due to the normalization in (25), Figure 3 suggests that quantity discounts do not fully explain demand synergies and the purchase of multiple units of CSDs.[46] Figure 3 also highlights that, in line with Figure 2, multi-person households have larger $\tilde{\gamma}_{t\mathbf{b}}^{hs}$, suggesting stronger preferences (or needs) for larger quantities of CSDs.[47] Finally, Figure 3 clarifies the complexity of specifying a priori restrictions on the demand synergies: even those corresponding to bundles with the same number of units of the same product are largely heterogeneous across markets and products.

To explore the determinants of demand synergies beyond quantity discounts, in Appendix Table 9 we report second-step OLS estimates of $\tilde{\gamma}_{t\mathbf{b}}^{hs} = \tilde{\Gamma}_{t\mathbf{b}}^{hs} + \alpha^{hs}\left(p_{t\mathbf{b}} - \sum_{j\in\mathbf{b}} p_{tj}\right)$ on observed bundle-level characteristics and fixed effects.[48] Net of quantity discounts, households appear to enjoy purchases of wider varieties of CSDs but also to dislike mixing products by different producers. Beyond quantity discounts, households like to purchase different CSDs (e.g., 1L of Coke and 1L of Sprite better than 2L 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 household sizes illustrates that single-person households have

[45] Relying instead on those from the other two columns of Table 3 leads to the same conclusions.

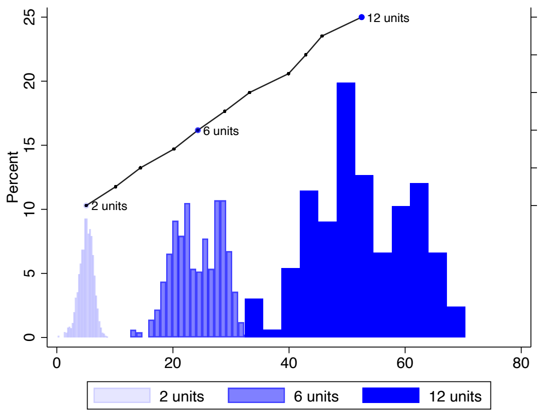
[46] Because of the normalization in (25), the fact that $\tilde{\gamma}_{t\mathbf{b}}^{hs}$ is on average large and that it increases almost linearly in $|\mathbf{b}|$ is consistent with the large observed share of shopping trips with no purchase of CSDs (76.29%, Table 1) and hence a large value of δ_{t0}^{hs}.

[47] Multi-person households have a smaller observed share of shopping trips with no purchase of CSDs than single person households (74.46% versus 82.71%), suggesting that $\delta_{t0}^{\text{multi}} < \delta_{t0}^{\text{single}}$. This in turn suggests that the estimated $\tilde{\gamma}_{t\mathbf{b}}^{\text{multi}} > \tilde{\gamma}_{t\mathbf{b}}^{\text{single}}$ should reflect $\gamma_{t\mathbf{b}}^{\text{multi}} > \gamma_{t\mathbf{b}}^{\text{single}}$.

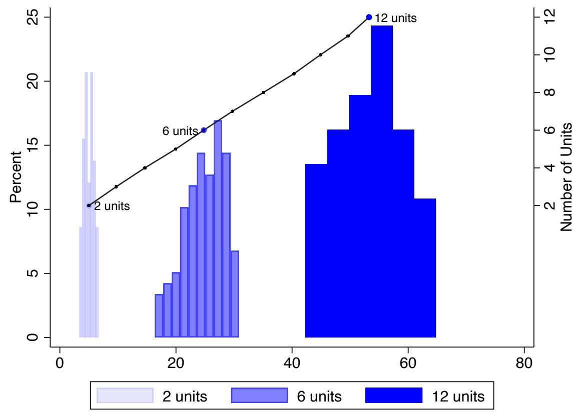
[48] As discussed in footnote 38, the C2SLS estimator cannot pin down the demand synergy $\Gamma_{t\mathbf{b}}$ of bundles that include products that are never observed to be purchased in isolation as single units in market t. In our PONL model, we have $81,215$ of these demand synergies. Because these cannot be part of the second-step OLS regression, the estimation sample used in Table 9 only includes $82,808$ of the $164,023$ total demand synergies. As detailed in Appendix G, 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.

Figure 3: Demand Synergies $\tilde{\gamma}^{hs}_{t\mathbf{b}}$ for bundles of 2, 6, and 12 Units

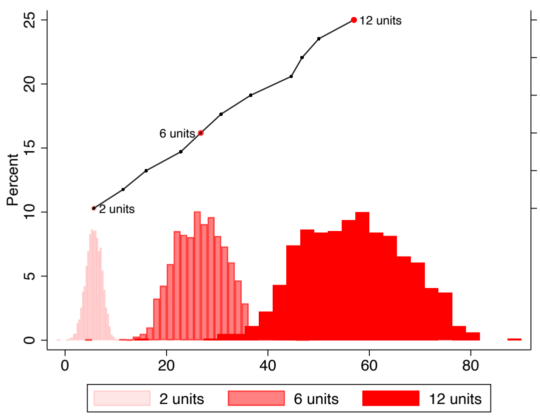
(a) Single-person households, same product



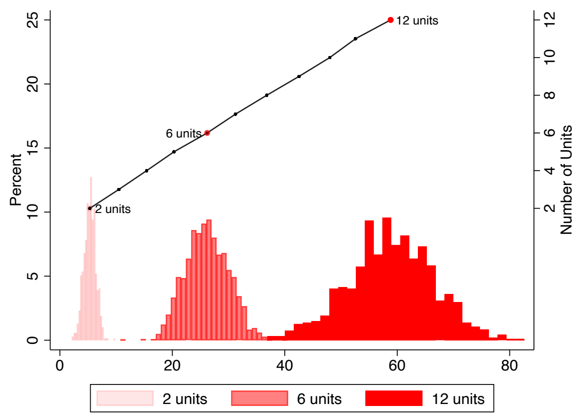
(b) Single-person households, different products



(c) Multi-person households, same product



(d) Multi-person households, different products



stronger preferences for this type of within-producer variety.

5.4 Estimated Elasticities

Our main objective is to evaluate the welfare effects of the observed quantity discounts by simulating a counterfactual with linear pricing (i.e., a constant unit-price for each product). As a way to summarize our estimation results and provide intuition for this counterfactual simulation, Table 4 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 $\sum_{j \in \mathbf{J}} \mathcal{s}^{hs}_{tj}$ and all multiple units $\sum_{\mathbf{b} \in \mathbf{C}_2} |\mathbf{b}| \times \mathcal{s}^{hs}_{t\mathbf{b}}$, 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

Table 4: Price Elasticities, by Household Size

	Single-person households		Multi-person households	
	$(p_{tj})_{j\in\mathbf{J}}$	$(p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$	$(p_{tj})_{j\in\mathbf{J}}$	$(p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$
Single units, $\sum_{j\in\mathbf{J}} s_{tj}^{\text{single}}$	-1.0543	1.1892	-1.5223	3.1967
	(0.1439)	(0.1759)	(0.1567)	(0.4489)
Multiple units, $\sum_{\mathbf{b}\in\mathbf{C}_2} \lvert\mathbf{b}\rvert s_{t\mathbf{b}}^{\text{single}}$	0.0197	-5.0360	0.0200	-7.0651
	(0.0031)	(0.6736)	(0.0017)	(0.5849)

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 H.1. Standard errors are computed using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

units $(p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$).[49] We derive the expressions of these price elasticities in Appendix H.1, while Table 4 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 4 focuses on price elasticities with respect to all prices of single units (first and third columns) and all prices of multiple units (second and fourth columns), as we expect these to be the most relevant when comparing quantity discounts to linear pricing. With quantity discounts, the producer of each product j sets all quantity-specific prices of j: price p_{tj} for a single unit of j, $p_{t(j,j)}$ for two units of j, and so on. 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 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 the 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 $(p_{tj})_{j\in\mathbf{J}}$.

Remember from Table 3 that single-person households are less price sensitive than multi-person households ($\alpha^{\text{single}} < \alpha^{\text{multi}}$). This implies the main patterns reported in Table 4: 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 -7.06% in the purchases of multiple units by multi-person households, but only of -5.04% 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 +3.2% increase in the purchases of single units by multi-person households, but only of +1.19%

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

by single-person households.

5.5 Counterfactual Simulation: Linear Pricing

To evaluate the welfare consequences of quantity discounts, we first rely on the PONL estimates from column (iii), Table 3, and calculate producers' marginal costs. 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. As detailed in Appendix H.2, we do this under the assumption that the observed prices were generated according to an oligopolistic Bertrand-Nash price-setting game of complete information that allows each product to have quantity-specific prices. Importantly, we do not assume producers to offer quantity discounts, but rather allow them to do so (along with the possibility of offering linear or even convex prices).

Table 5 summarizes our marginal cost estimates among bundles of a single or multiple units of the same product, $\mathbf{b} = (j, ..., j)$, and across markets. We regress the estimated marginal costs per unit on the number of units in the bundle and various fixed effects.[50] On average, estimated marginal costs per unit increase with quantity: selling multiple units of CSDs jointly is, if anything, more costly for producers than selling them separately.[51] This and Figure 3 suggest that producers' incentives to offer quantity discounts are mainly driven by demand synergies rather than by cost savings.

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 H.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 H.4. Tables 6 and 7 summarize these results in terms of median changes across the same set of markets used in Table 4. As for the price elasticities, standard errors are computed 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 -43.8 cents and to a

[50]The regression includes the marginal costs estimated to be between the 1th-99th percentiles. The results are robust to including the estimated marginal costs between the 3th-97th or between the 5th-95th percentiles.

[51]Our estimates include *all* variable costs incurred to sell CSDs to households, including both the costs of production and those of retailing. In this sense, handling larger volumes of CSDs may lead to larger transportation costs, stocking costs, and opportunity costs of shelf space at the point of sale in retailers.

Table 5: Summary of Estimated Marginal Costs

	Marginal Cost per Unit ($ per liter)	
Number of Units		
Overall	0.066 (0.003)	
Coca-Cola		0.059 (0.004)
PepsiCo		0.102 (0.008)
Dr. Pepper		0.043 (0.003)
Others		0.026 (0.009)
Product fixed effects	yes	yes
Store fixed effects	yes	yes
Time fixed effects	yes	yes
Num. of Obs.	$118,357$	$118,357$
R^2	0.092	0.092

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

	Average	Single-person households	Multi-person households
Price Change ($)			
Δp_{tj}	-0.438 (0.0162)		
$\Delta p_{t\mathbf{b}}$	$+0.894$ (0.0703)		
Quantity Change (L per household-year)	-32.136 (2.027)	-12.766 (0.696)	-35.303 (2.120)
Single units	$+0.473$ (0.021)	$+0.268$ (0.027)	$+0.517$ (0.023)
Multiple units	-32.624 (2.056)	-13.163 (0.712)	-35.594 (2.200)
Quantity Change (%)	-20.66% (1.51%)	-18.99% (1.41%)	-21.26% (1.60%)
Conditional on purchase	-11.94% (0.85%)	-9.52% (1.07%)	-12.02% (1.03%)
ΔProb. of purchasing	-8.73% (0.78%)	-8.19% (0.81%)	-9.14% (0.78%)

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

simultaneous increase in the prices of multiple units of +89.4 cents. With respect to observed quantity discounts, these price changes are substantial and correspond to a decrease of −31.62% and to an increase of +14.91%, respectively. Intuitively, these price changes are expected to make purchases of smaller quantities relatively more convenient for both household sizes, as confirmed by the middle panel of Table 6: yearly purchased quantities per household would decrease by −32.14 liters, obtained as the difference between a small increase in purchases of single units (+0.47 liters) and a large reduction in purchases of multiple units (−32.62 liters).[52] The bottom panel of Table 6 shows that these large reductions in purchased quantities (−20.66%) are motivated by both a substitution from purchases

[52]To avoid problems with outliers, each entry of Tables 6 and 7 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 H.4 for details.

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

	Average	Single-person households	Multi-person households
Profit Change ($ per household-year)	−7.466	−3.510	−8.390
	(0.955)	(0.551)	(0.969)
Single units	+0.028	−0.051	+0.058
	(0.017)	(0.018)	(0.021)
Multiple units	−7.622	−3.510	−8.580
	(0.953)	(0.549)	(0.964)
CV ($ per household-year)	+3.698	+1.752	+4.119
	(0.203)	(0.227)	(0.247)
Single units	−0.403	−0.364	−0.394
	(0.075)	(0.067)	(0.086)
Multiple units	+4.242	+2.256	+4.592
	(0.179)	(0.195)	(0.209)
CV/Expenditure (% per household-year)	+2.82%	+3.11%	+2.68%
	(0.13%)	(0.30%)	(0.14%)

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

of larger quantities toward purchases of smaller quantities (−11.94%) and by a decrease in the probability of purchasing CSDs altogether (−8.73%).

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 2 and the price elasticities in Table 4. While in relative terms multi-person households would reduce their purchased quantities only around 2.3 percentage points more than single-person households (−21.26% versus −18.99%, 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 by −35.3 liters per year, almost three times more than single-person households (−12.77 liters per year). The vast majority of this difference stems from the larger reduction in purchases of multiple units by multi-person relative to single-person households (−35.59 versus −13.16 liters per year). This can be explained by noting that multi-person households both have higher price elasticity of demand for multiple units (−7.06% versus −5.03%, Table 4) and purchase multiple units in greater amounts under quantity discounts (Figure 2).

The top panel of Table 7 illustrates that this striking reduction in purchased quantities of −20.66% would cause a decrease in yearly profit per household of −7.47$ (−19.74%), obtained as the difference between a very small per household-year profit increase from purchases of single units (+2.8 cents) and a very large per household-year profit reduction from purchases of multiple units (−7.62$). 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 (−8.39\$, or −21.02%, versus −3.51\$, or −15.53%), 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 +3.7\$ per household-year would be necessary for households to remain indifferent between quantity discounts and linear pricing, corresponding to 2.82% 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: +4.12\$ per household-year (2.68% of expenditure) as opposed to +1.75\$ (+3.11% of expenditure). As discussed in Appendix H.5, these 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 lose more by linear pricing.

6 Reduction in Added Sugar Intake

From the above counterfactual simulations, we can draw some conclusions about a ban on quantity discounts. First, quantity discounts seem profitable for producers of CSDs in the USA and this is not motivated by cost savings but rather by the demand synergies associated to purchases of larger quantities. 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).

Second, despite the substantial reduction in quantity purchased (−20.66%), consumer surplus would not reduce too sharply, with a compensating variation of +3.7\$ per household-year (amounting to 2.82% 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.

These observations suggest that a ban on quantity discounts as the one proposed on 6 April 2022 by the UK government (see Introduction) could serve as a practically simple and effective 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. A ban on quantity discounts would lead households to drastically reduce the purchases 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, we investigate this possibility and simulate the reduction in added sugar intake from CSDs implied by (i) a *total* ban on quantity discounts on all CSDs (as in section 5.5) and (ii) a *partial* ban on quantity discounts only on the CSDs with added sugar. To put these results in perspective, we also calibrate (iii) the amount of sugar tax per ounce of added sugar that would generate similar reductions in added sugar intake as (ii).[53] 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:[54] 50% of the CSDs in our analysis have added sugar (sugary CSDs), while the remaining 50% do not (non-sugary CSDs). Households are observed to purchase an average of 60L a year of sugary and 57.24L of non-sugary CSDs.

The first and second columns of Table 8 summarize counterfactual results for (i) and (ii) (see details in Appendix H.4). With linear pricing, households would reduce their yearly purchased quantities of added sugar from CSDs by -22.93%, a similar reduction as that implied by a ban on quantity discount for sugary CSDs (-22.08%). However, while linear pricing would lead to a large reduction in the purchased quantities of *all* CSDs (-20.66%), the ban on quantity discounts for sugary CSDs would instead lead to a large reduction in the purchased quantities of sugary CSDs (around -21.89%) with some substitution to non-sugary CSDs (around $+1.71\%$). These different purchase patterns would give rise to different welfare implications, with the partial ban on quantity discounts leading to remarkably smaller profit losses (around -9.46% instead of -19.74%) and compensating variations (around $+1.77\$$ instead of $+3.7\$$), suggesting that, in terms of reducing added sugar intake, a targeted ban on quantity discounts for sugary CSDs may be more efficient than linear pricing for all CSDs.

The third column of Table 8 reports our calibration results for (iii), a sugar tax that would approximately generate similar reductions in added sugar intake

[53]Importantly, the cities of Pittsfield and Eau Claire were not subject to a sugar tax on CSDs in the period 2008-2011.

[54]These additional data on added sugar content are available from the authors on request.

Table 8: Counterfactual Linear Pricing: Changes in Added Sugar

	Ban on quantity discounts on all CSDs	only on sugary CSDs	Sugar tax 1¢/oz of added sugar
Predicted added sugar change	−22.93%	−22.08%	−22.90%
Quantity change	−20.66%	−8.93%	−10.35%
Sugary CSDs	−23.95%	−21.89%	−21.98%
Non-Sugary CSDs	−17.83%	+1.71%	+2.14%
Profit change	−19.74%	−9.46%	−7.01%
CV (per household-year)	+3.70$	+1.77$	+2.35$
CV/Expenditure	+2.82%	+1.29%	+1.61%

Notes: The first column reports the changes due to a ban on quantity discounts on all CSDs (as in Tables 6 and 7), the second due to a ban on quantity discounts only on the sugary CSDs, while the third due to a sugar tax of 1¢/oz of added sugar. The above changes are obtained in a similar manner to the counterfactual quantity changes in Tables 6 and 7, see computational details in Appendix H.4. All entries are computed as medians over the same set of markets used in Table 4 to compute price elasticities. Standard errors are obtained using the parametric bootstrap procedure described in Appendix B.1 with 200 repetitions.

as (ii) (see details in Appendix H.4). These results suggests that—also in terms of changes in purchase patterns, reductions in profits, and compensating variations—a ban on quantity discounts for sugary CSDs can be thought of as a sugar tax of 1 cent per ounce of added sugar. Beyond these calibrations, the existing literature also confirms that the reduction in intake of added sugar from CSDs we find in our simulations is of a similar order as that obtained by typical sugar taxes.[55]

Further research should investigate the many important dimensions of comparison with sugar taxes we did not discuss, such as potential regressivity, internalities and externalities of sugar intake, and redistribution of tax revenue. However, our results suggest that a ban on quantity discounts could be an effective and easily implementable policy to limit the intake of added sugar from CSDs. While a ban on quantity discounts can be implemented by enforcing linear pricing on all or only some CSDs, the effective design and implementation of sugar taxes rely on information not always available, such as measures of the externalities and the internalities of sugar intake (Allcott et al., 2019; O'Connell and Smith, 2020), and on a more involved participation of the government 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," *Quarterly Journal of Economics*, 90(3), 475–98.

[55]See footnote 7 for a summary of this literature.

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.

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.

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.

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.

BRONNENBERG, B. J., M. W. KRUGER, AND C. F. MELA (2008): "Database paper—The IRI marketing data set," *Marketing science*, 27(4), 745–748.

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.

Chesher, A., and J. M. Santos Silva (2002): "Taste variation in discrete choice models," *The Review of Economic Studies*, 69(1), 147–168.

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. S., R. Griffith, and A. Iaria (2021): "A survey of preference estimation with unobserved choice set heterogeneity," *Journal of Econometrics*, 222(1), 4–43.

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," *The American Economic Review*, 102(2), 643–685.

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

Grieco, P. L., C. Murry, J. Pinkse, and S. Sagl (2022): "Conformant and Efficient Estimation of Discrete Choice Demand Models," .

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.

Hallett, A. H. (1982): "Alternative techniques for solving systems of nonlinear equations," *Journal of Computational and Applied Mathematics*, 8(1), 35–48.

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 (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.

——— (2022): "The Mixed Logit and Mixed Probit are Real Analytic," *Available at SSRN 4094866*, pp. 1–16.

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

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

Lanier, J., J. Large, and J. Quah (2022): “Estimating very large demand systems,” *Working Paper*, pp. 1–62.

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*.

Mugnier, M., and A. Wang (2022): “Identification and (Fast) Estimation of Large Nonlinear Panel Models with Two-Way Fixed Effects,” *Available at SSRN 4186349*.

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.*

Papola, A. (2004): "Some developments on the cross-nested logit model," *Transportation Research Part B: Methodological*, 38(9), 833–851.

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.

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.

Salanié, B., and F. A. Wolak (2019): "Fast," robust", and approximately correct: estimating mixed demand systems," Discussion paper, National Bureau of Economic Research.

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.*

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.

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]

November, 2022

Appendix

A Proofs

A.1 Proof of Proposition 1

Without loss of generality, suppose that $\mathbf{N}_j$ includes at least another bundle in addition to a single unit of j for $j = 1, ..., J$.[56]

Assumption 1.

(i) 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 $\{(\mathscr{s}_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C_1}} : \sum_{\mathbf{b}\in\mathbf{C}_1} \mathscr{s}_{t\mathbf{b}} < 1, \mathscr{s}_{t\mathbf{b}} > 0, \mathbf{b} \in \mathbf{C}_1\}$.

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

Assumption 1(i) 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(ii) is a standard completeness condition in the literature on identification of demand using instrumental variables (Berry and Haile, 2014).

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

Proof. 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 ∞, the left- and the right-hand sides will cross only once and (16) will have a unique solution. $\square$

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

[56]Otherwise, purchase probabilities (7) would not depend on λ_j.

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

$$\pi_{tj} = \left[\textstyle\sum_{\mathbf{b}' \in \mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j} \exp\left(\Gamma_{\mathbf{b}'}(\lambda, \pi_t; \mathscr{s}_t)/\lambda_j\right) \prod_{r \in \mathbf{b}'} \left[\frac{\mathscr{s}_{tr}}{\mathscr{s}_{t0}}\right]^{\lambda_r/\lambda_j} \pi_{tr}^{(1-\lambda_r)/\lambda_j}\right]^{\lambda_j}, j = 1, ..., J. \quad (26)$$

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 of (26) when $\lambda = \lambda^0$. Denote $\pi_t^0 = \pi^0(\lambda^0; \mathscr{s}_t)$. Then, at $(\alpha^0, \beta^0, \delta^0, \lambda^0)$, for $j = 1, ..., J$,

$$\begin{aligned} &\ln \mathscr{s}_{tj} - \ln \mathscr{s}_{t0} = \delta_j^0 + x_{tj}\beta^0 - \alpha^0 p_{tj} + (1 - \lambda_j^0) \ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}^0}\right) + \xi_{tj}, \\ &\mathbb{E}\left[g_j^0(\mathscr{s}_t, p_{tj}, x_{tj}; \alpha^0, \beta^0, \delta^0, \lambda^0) | Z_t = Z\right] = 0, \end{aligned} \quad (27)$$

where

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

Using Assumption 1(ii), we identify $g_j^0 = g_j^0(\mathscr{s}_t, p_{tj}, x_{tj}; \alpha^0, \beta^0, \delta^0, \lambda^0)$ as a function of $(\mathscr{s}_t, p_{tj}, x_{tj})$ for each $j = 1, ..., J$. Then, due to Assumption 1(i), 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 $\mathscr{s}_{tj} \to 1$ and therefore $\ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_j^0(\lambda^0; \mathscr{s}_t)}\right) = \ln\left(\mathscr{s}_{tj}/\mathscr{s}_t^j\right) \to 0$, we identify δ_j^0. As a result, we identify the quantities $A_{tj}^0 = (1-\lambda_j^0) \ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_j^0(\lambda^0; \mathscr{s}_t)}\right) = (1-\lambda_j^0) \ln\left(\mathscr{s}_{tj}/\mathscr{s}_t^j\right)$ for $t \in \mathbf{T}$ and $j \in \mathbf{J}$. Using $\sum_{j=1}^J \pi_{tj}^0 + 1 = 1/\mathscr{s}_{t0}$, we obtain that for each $t \in \mathbf{T}$, $\lambda = \lambda^0$ satisfies:

$$\sum_{j=1}^J \frac{\mathscr{s}_{tj}}{1 - \mathscr{s}_{t0}} \exp\left(\frac{-A_{tj}^0}{1 - \lambda_j}\right) = 1 \quad (28)$$

or equivalently

$$\sum_{j=1}^J \frac{1}{1 - \mathscr{s}_{t0}} \left(\mathscr{s}_t^j\right)^{\frac{1-\lambda_j^0}{1-\lambda_j}} \mathscr{s}_{tj}^{\frac{\lambda_j^0 - \lambda_j}{1-\lambda_j}} = 1. \quad (29)$$

We now show that λ^0 is the only $\lambda \in \mathbb{R}^J$ that satisfies (29) and therefore identified.

Because of Assumption 1(ii), for each $j = 1, ..., J$, we can keep all $\mathscr{s}_{t\mathbf{b}}$, $\mathbf{b} \in \mathbf{N}_j$ and $\mathbf{b} \neq j$ constant and positive, while let $\mathscr{s}_{tj} \to 0$. Note that $\frac{1}{1-\mathscr{s}_{t0}} \left(\mathscr{s}_t^j\right)^{\frac{1-\lambda_j^0}{1-\lambda_j}}$ is always bounded away from zero and bounded from above for all $j = 1, ..., J$; in contrast,

$\mathscr{s}_{tj}^{\frac{\lambda_j^0-\lambda_j}{1-\lambda_j}}$ tends to $+\infty$ if $\lambda_j^0 < \lambda_j$. Then, for (29) to hold, $\lambda_j^0 \geq \lambda_j$ for $j = 1, ..., J$. Also note that given $\mathscr{s}_t$ and $\{A_{tj}^0\}_{j=1}^J$, the left-hand side of (28) is strictly increasing with respect to λ_j for $j = 1, ..., J$. Then, the only feasible λ satisfying $\lambda_j^0 \geq \lambda_j$ and (28) is $\lambda = \lambda^0$. Finally, using Lemma 1, we then identify all $\Gamma_{t\mathbf{b}}$'s.

B Proof of Proposition 2

Throughout the proof, we use $|\cdot|$ to refer to the Euclidean norm. 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 $j = 1, ..., J$:

$$\pi_{tj} = \phi_j(\pi_t; \lambda, \mathscr{s}_t) = \left[\sum_{\mathbf{b}'\in\mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j} \exp\left(\Gamma_{\mathbf{b}'}(\lambda, \pi_t; \mathscr{s}_t)/\lambda_j\right) \prod_{r\in\mathbf{b}'} \left[\frac{\mathscr{s}_{tr}}{\mathscr{s}_{t0}}\right]^{\lambda_r/\lambda_j} \pi_{tr}^{(1-\lambda_r)/\lambda_j}\right]^{\lambda_j}. \quad (30)$$

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*

$$\inf_{t\in\mathbf{T}, |\lambda-\lambda^0|\leq a, |\pi_t-\pi_t^0|\leq b} \left| Det\left(\frac{\partial\Phi(\pi_t; \lambda, \mathscr{s}_t)}{\partial\pi}\right)\right| > \eta, \quad (31)$$

and

$$\sup_{t\in\mathbf{T}, |\lambda-\lambda^0|\leq a, |\pi_t-\pi_t^0|\leq b} \left| Det\left(\frac{\partial\Phi(\pi_t; \lambda, \mathscr{s}_t)}{\partial\lambda}\right)\right| \leq M, \quad (32)$$

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 $t \in \mathbf{T}$.

Assumption 2 summarizes the regularity conditions needed for Proposition 2. Condition (31) is a rank condition of non-linear system (30) with respect to π_t, uniformly for all $t \in \mathbf{T}$. It guarantees that each π_t can be expressed as a function of λ given $\mathscr{s}_t$ in a neighborhood of λ^0. Condition (32) requires that the Jacobian of this function from λ to π_t is uniformly bounded in a 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 $t \in \mathbf{T}$.[57] 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 $\mathscr{s}_t$: $\Gamma_t = \Gamma(\lambda, \pi(\lambda; \mathscr{s}_t); \mathscr{s}_t)$.

[57] 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 $t \in \mathbf{T}$.

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

$$\begin{aligned}\theta = (\delta, \beta, \alpha, 1-\lambda)^{\mathrm{T}} &= \psi_T(\theta) \\ &= \left(X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})X(\theta)\right)^{-1} X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})Y\end{aligned} \tag{33}$$

where

$$X(\theta) = (x_{tk}(\theta))_{t=1,...,T;k=1,...,2J+K+1} = \left((e_j)_{j=1}^J, x_t, -p_t, \left(\ln\left(\frac{\mathcal{s}_{tj}/\mathcal{s}_{t0}}{\pi_j(\lambda; \mathcal{s}_t)}\right)\right)_{j=1}^J\right)_{t=1}^T,$$

$$Y = (y_t)_{t=1,...,T} = (\ln(\mathcal{s}_{t1}/\mathcal{s}_{t0}, ..., \mathcal{s}_{tJ}/\mathcal{s}_{t0}))_{t=1,...,T} \in \mathbb{R}^{JT\times 1}.$$

Step 2: (Finite sample) Existence of a solution to (33)**.** We now prove that when T is large enough, (33) has a solution in a fixed neighborhood of θ^0 (i.e., the neighborhood does not depend on T). Define

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

For a vector-value function, we use superscript l to refer to its l^{th} component.

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 $v \in \mathbb{R}^{2J+K+1}$, $l = 1, ..., 2J+K+1$,

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

(ii) Det$(G_{XZ}(\theta)G_{XZ}(\theta)^T) > 0$ uniformly on $|\theta - \theta^0| \leq d_0$.

(iii) For $l = 1, ..., 2J+K+1$,

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

Because of Assumptions 3(ii) and (iii), $\frac{\partial^k\psi_T(\theta)}{\partial\theta^k}$ converges uniformly to $\frac{\partial^k\psi(\theta)}{\partial\theta^k}$ in $|\theta - \theta^0| \leq d_0$ with probability one. Then, combining this with Assumption 3(i), we obtain that there exists $M_1 > \frac{2J+K+1}{r} > 0$ such that

$$\left| \left(\frac{\psi_T(\theta)}{\partial \theta} - \mathbf{I} \right)^{-1} v \right| \leq M_1 |v|, \quad \left| \frac{\partial^2 \psi_T^{(l)}(\theta)}{\partial \theta^2} v \right| \leq M_1 |v| \tag{34}$$

uniformly for $l = 1, ..., 2J+K+1$, $v \in \mathbb{R}^{2J+K+1}$, and $|\theta - \theta^0| \leq d_0$ with probability one as $T \to \infty$. Now consider the following Newton-Raphson procedure:

$$\begin{aligned} \theta_0 &= \theta^0, \\ \theta_{k+1} &= -\left[\frac{\partial \psi_T(\theta_k)}{\partial \theta} - \mathbf{I} \right]^{-1} (\psi_T(\theta_k) - \theta_k) + \theta_k. \end{aligned} \tag{35}$$

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 $\jmath_{t(j|j)}$, which we denote by θ_T^{2SLS}. Note that $\psi_T(\theta_0) = \theta_T^{\text{2SLS}} \xrightarrow{p} \theta^0$ as $T \to \infty$.

Lemma 2. *Suppose that Assumptions 2-3 hold. In addition, $|\theta_T^{2SLS} - \theta^0| \leq \epsilon$, where $\epsilon > 0$ and $\epsilon \times M_1 < d_0/2$, with M_1 being defined in* (34) *and $d_0 \in (0,1)$. Then, for any $k > 0$, we have $|\theta_k - \theta_{k-1}| \leq \left(\frac{d_0}{2}\right)^k$ 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

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

Then, a 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 (35), and Assumption 3 imply:

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

where $r_2(\theta_1 - \theta_0) = \left((\theta_1 - \theta_0)^{\mathrm{T}} \frac{\partial^2 \psi^{(l)}(\tilde{\theta}^{(l)})}{\partial \theta^2} (\theta_1 - \theta_0) \right)_{l=1}^{2J+K+1}$ with $\tilde{\theta}^{(l)}$ between θ_1 and θ_0. 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 and (35):

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

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

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

The proof is complete. □

Note that the event that (34) and $|\theta_T^{2\text{SLS}} - \theta^0| \le \epsilon$ jointly hold occurs with probability one as $T \to \infty$. Because $d_0 \in (0,1)$, then Lemma 2 implies that with probability one: (1) θ_k converges to some θ^* such that $|\theta^* - \theta^0| \le d_0$ and (2) $\psi_T(\theta^*) = \theta^*$, i.e. the existence of a solution to (33). 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 (33), we can re-formulate $\hat{\theta}$ as an extremum estimator:

$$
\begin{aligned}
\hat{\theta} &= \underset{\theta:|\theta-\theta^0|\le d_0}{\operatorname{argmin}}\ Q_T(\theta), \\
Q_T(\theta) &= |\theta - \left(X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})X(\theta)\right)^{-1} X^{\mathrm{T}}(\theta)(ZZ^{\mathrm{T}})Y|^2.
\end{aligned}
\tag{36}
$$

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 $\{\theta : |\theta - \theta^0| \le d_0\}$. Moreover, due to the definition of $x_{tk}(\theta)$ and Assumption 3, $Q(\theta)$ is continuous. Finally, because of Assumption 3(iii), $X^{\mathrm{T}}(\theta)Z/T \xrightarrow{p} G_{XZ}(\theta)$ uniformly for θ in $|\theta - \theta^0| \le d_0$. Then, $Q_T(\theta) \xrightarrow{p} Q(\theta)$ uniformly for θ in $|\theta - \theta^0| \le d_0$. 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 (33) at $\theta = \hat{\theta}$ around $\theta = \theta^0$:

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

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

$$\sqrt{T}\left(\hat{\theta} - \theta^0\right) = \left[\mathbf{I} - \frac{\partial \psi_T(\tilde{\theta})}{\partial \theta}\right]^{-1} \sqrt{T}\left(\theta_T^{\text{2SLS}} - \theta^0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma V^{\text{2SLS}} \Sigma^{\mathrm{T}}), \tag{37}$$

where V^{2SLS} is the asymptotic variance-covariance matrix of θ_T^{2SLS} and

$$\Sigma = \left[\mathbf{I} - \frac{\partial \left[\left(G_{XZ}(\theta^0) G_{XZ}^{\mathrm{T}}(\theta^0)\right)^{-1} G_{XZ}(\theta^0) G_{YZ}\right]}{\partial \theta}\right]^{-1}. \tag{38}$$

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 $\mathcal{s}_t$) 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 on objects that we derive from these parameters in our counterfactuals.

Inference on θ. We provide consistent estimators of V^{2SLS} and Σ in (37). Given the consistency of $\hat{\theta}$, a plug-in estimator of V^{2SLS} is:

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

where $\hat{G}_{XZ}(\hat{\theta}) = \left(\frac{\sum_{t=1}^{T} x_{tk}(\hat{\theta}) z_{tk'}}{T}\right)_{k=1,\ldots,2J+K+1, k'=1,\ldots,P} \in \mathbb{R}^{(2J+K+1) \times P}$ 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 $\hat{\pi}_t$, $t = 1, \ldots, T$.

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}:

$$
\begin{aligned}
&\frac{\partial\left[\left(G_{XZ}G_{XZ}^{\mathrm{T}}\right)^{-1}G_{XZ}G_{YZ}\right]}{\partial\theta}\\
&=\left[\frac{\partial\left[\left(G_{XZ}G_{XZ}^{\mathrm{T}}\right)^{-1}\right]}{\partial\theta}G_{XZ}+\left(G_{XZ}G_{XZ}^{\mathrm{T}}\right)^{-1}\frac{\partial G_{XZ}}{\partial\theta}\right]G_{YZ} \qquad (39)\\
&=\left(G_{XZ}G_{XZ}^{\mathrm{T}}\right)^{-1}\left[\frac{\partial G_{XZ}}{\partial\theta}-\frac{\partial G_{XZ}}{\partial\theta}G_{XZ}^{\mathrm{T}}(G_{XZ}G_{XZ}^{\mathrm{T}})^{-1}G_{XZ}-G_{XZ}\frac{\partial G_{XZ}^{\mathrm{T}}}{\partial\theta}(G_{XZ}G_{XZ}^{\mathrm{T}})^{-1}G_{XZ}\right]G_{YZ}.
\end{aligned}
$$

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 (39) to obtain a consistent estimator of $\frac{\partial\left[\left(G_{XZ}G_{XZ}^{\mathrm{T}}\right)^{-1}G_{XZ}G_{YZ}\right]}{\partial\theta}$. Finally, we plug this consistent estimator in (38) 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 its explicit formula from (30).

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.[58] 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 the asymptotic distribution of $\hat{\theta}$ in (37). 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 $B=200$.

[58]The iterative procedure stops when the non-linear 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.

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 these.

C Details on the Iterative Estimation Procedure

C.1 Iteration 0: Choice of Starting Values

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 the 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 $\pi_{tj}^{(0)} = \frac{\sum_{\mathbf{b}\in\mathbf{N}_j} \omega_{\mathbf{b}j} s_{t\mathbf{b}}}{s_{t0}}$, replacing each unobserved within-nest market share $s_{t(\mathbf{b}|j)}$ by its corresponding 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 non-linear equations (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):

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

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 non-linear equations (16) and (17).

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

$$G_{t\mathbf{b}}^{(k-1)} = \exp\left(\frac{\Gamma_{t\mathbf{b}}^{(k-1)}\left(1-\lambda_j^{(k)}\right)}{\lambda_j^{(k)}}\right) \left(\omega_{\mathbf{b}j}\right)^{\frac{1}{\lambda_j^{(k)}}} \left(\pi_{tj}^{(k)}\right)^{1-\frac{1}{\lambda_j^{(k)}}} \prod_{r\in\mathbf{b}} \left[\mathscr{s}_{tr}/\mathscr{s}_{t0}\right]^{\frac{\lambda_r^{(k)}\left(1-\lambda_j^{(k)}\right)}{\lambda_j^{(k)}}} \left(\pi_{tr}^{(k)}\right)^{\frac{\left(1-\lambda_r^{(k)}\right)\left(1-\lambda_j^{(k)}\right)}{\lambda_j^{(k)}}} \tag{40}$$

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

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

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 $k = \bar{K}$, exit the algorithm.

As mentioned in footnote 36, we define numerical convergence—and hence the choice of $\bar{K}$—on the basis of 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 Appendix F, our Monte Carlo simulations suggest that $\bar{K} = 5$ iterations can already be sufficient to achieve this form of numerical convergence.

D Proof of Proposition 3

Our iterative estimation procedure is:

$$\begin{aligned}
X^{(k)} &= \left(\left(e_j\right)_{j=1}^{J}, x_t, -p_t, \left(\ln\left(\frac{\mathscr{s}_{tj}/\mathscr{s}_{t0}}{\pi_{tj}^{(k)}}\right)\right)_{j=1}^{J}\right)_{t=1}^{T}, \\
(\alpha^{(k+1)}, \beta^{(k+1)}, \delta^{(k+1)}, 1-\lambda^{(k+1)})^{\mathrm{T}} &= \left(X^{(k)\mathrm{T}}(ZZ^{\mathrm{T}})X^{(k)}\right)^{-1}\left(X^{(k)\mathrm{T}}(ZZ^{\mathrm{T}})Y\right), \\
\Gamma_{t\mathbf{b}}^{(k+1)} &= \Gamma_{t\mathbf{b}}^{(k)} - \left(\Gamma_{t\mathbf{b}}^{(k)} - \left(\ln[\mathscr{s}_{t\mathbf{b}}/\mathscr{s}_{t0}] - \sum_{j\in\mathbf{b}}\left(\lambda_j^{(k+1)}\ln[\mathscr{s}_{tj}/\mathscr{s}_{t0}] + \left(1-\lambda_j^{(k+1)}\right)\ln\pi_{tj}^{(k)}\right)\right) + \ln\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k)}\right) \\
&\quad \times \frac{\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k)}}{\sum_{j=1}^{J} G_{t\mathbf{b}}^{(k)}/\lambda_j^{(k+1)}} \\
\pi_{tj}^{(k+1)} &= \left[\sum_{\mathbf{b}'\in\mathbf{N}_j} \omega_{\mathbf{b}'j}^{1/\lambda_j^{(k+1)}} \exp\left(\Gamma_{\mathbf{b}'}^{(k+1)}/\lambda_j^{(k+1)}\right) \prod_{r\in\mathbf{b}'}\left[\frac{\mathscr{s}_{tr}}{\mathscr{s}_{t0}}\right]^{\lambda_r^{(k+1)}/\lambda_j^{(k+1)}} \left[\pi_{tr}^{(k)}\right]^{(1-\lambda_r^{(k+1)})/\lambda_j^{(k+1)}}\right]^{\lambda_j^{(k+1)}},
\end{aligned}$$

where $G_{t\mathbf{b}}^{(k)}$ is defined in (40). Because $\pi_t^{(k)} \to \pi_t^*$ for all $t \in \mathbf{T}$, then $X^{(k)}$ and $(\alpha^{(k)}, \beta^{(k)}, \delta^{(k)}, 1-\lambda^{(k)})$ 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:

$$\Gamma_{t\mathbf{b}}^{*} = \ln[\mathcal{s}_{t\mathbf{b}}/\mathcal{s}_{t0}] - \sum_{j\in\mathbf{b}} \left(\lambda_j^{*} \ln[\mathcal{s}_{tj}/\mathcal{s}_{t0}] + \left(1-\lambda_j^{*}\right) \ln \pi_{tj}^{*}\right)$$
$$- \ln\left(\sum_{j=1}^{J} \exp\left(\frac{\Gamma_{t\mathbf{b}}^{*}(1-\lambda_j^{*})}{\lambda_j^{*}}\right) (\omega_{\mathbf{b}j})^{\frac{1}{\lambda_j^{*}}} \left(\pi_{tj}^{*}\right)^{1-\frac{1}{\lambda_j^{*}}} \prod_{r\in\mathbf{b}} [\mathcal{s}_{tr}/\mathcal{s}_{t0}]^{\frac{\lambda_r^{*}(1-\lambda_j^{*})}{\lambda_j^{*}}} \left(\pi_{tr}^{*}\right)^{\frac{(1-\lambda_r^{*})(1-\lambda_j^{*})}{\lambda_j^{*}}}\right).$$

Consequently, at $(\alpha^*, \beta^*, \delta^*, 1-\lambda^*)$, $\Gamma_{t\mathbf{b}}^{*}$, and π_t^*, equations (16) are satisfied. Similarly, equations (17) are satisfied. Therefore, $(\delta^*, \beta^*, \alpha^*, \lambda^*)$ and $(\pi_t^*, \Gamma_t^*)_{t\in\mathbf{T}}$ satisfy (22).

E Extension: Adding a Random Coefficient

We consider a PONL with a centered and normalized random coefficient ν_i, distributed according to F, that multiplies price $p_{t\mathbf{b}}$:[59]

$$U_{it\mathbf{b}}(\sigma) = \sigma\nu_i p_{t\mathbf{b}} + \delta_{t\mathbf{b}} + \varepsilon_{it\mathbf{b}}. \tag{41}$$

Suppose that ν_i is independent of $(\varepsilon_{it\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_1\cup\{0\}}$. Then, the purchase probability of bundle $\mathbf{b}$ is:

$$\mathcal{s}_{t\mathbf{b}} = s_{\mathbf{b}}(\delta_t; \sigma) = \int s_{\mathbf{b}}(\sigma v_i p_t + \delta_t; 0) dF(v_i), \tag{42}$$

where $s_{\mathbf{b}}(\cdot; 0)$ refers to the purchase probability of $\mathbf{b}$ in PONL model (7).

Sketch of Identification, Estimation, and Implementation. System $(s_{\mathbf{b}}(\delta_t; 0))_{\mathbf{b}\in\mathbf{C}_1}$ is a PONL model without random coefficients satisfying:

$$\begin{aligned} \ln\frac{s_j(\delta_t;0)}{s_0(\delta_t;0)} &= \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j)\ln\left(\frac{s_j(\delta_t;0)/s_0(\delta_t;0)}{\pi_{tj}}\right) + \xi_{tj} \\ \Gamma_t &= (\Gamma_{\mathbf{b}}(\Gamma_{t\mathbf{b}}; \lambda, \pi_t, s(\delta_t;0)/s_0(\delta_t;0)))_{\mathbf{b}\in\mathbf{C}_2} \text{ from (16)} \\ \pi_t &= \left(\phi_j(\pi_t; \lambda, \Gamma_t, s(\delta_t;0)/s_0(\delta_t;0))\right)_{j\in\mathbf{J}} \text{ from (17).} \end{aligned} \tag{43}$$

If we knew $(s_{\mathbf{b}}(\delta_t; 0))_{\mathbf{b}\in\mathbf{C}_1}$ (and therefore $s_0(\delta_t; 0)$), we could then identify and estimate all the parameters except σ as in the standard PONL model (7). Identification and estimation in the presence of random coefficients consists in first recovering $(s_{\mathbf{b}}(\delta_t; 0))_{\mathbf{b}\in\mathbf{C}_1}$ from the observed purchase probabilities and then σ.

To this aim, we rely on the Taylor expansion of $\ln\frac{\mathcal{s}_{t\mathbf{b}}}{\mathcal{s}_{t0}} = \ln\frac{s_{t\mathbf{b}}(\sigma)}{s_{t0}(\sigma)}$ with respect

[59]Similarly, we could also incorporate random coefficients that capture unobserved heterogeneity in other dimensions, such as bundle size.

to σ around $\sigma = 0$ for each $\mathbf{b} \in \mathbf{C}_1$:

$$\ln \frac{\mathscr{s}_{t\mathbf{b}}}{\mathscr{s}_{t0}} = \ln \frac{s_{\mathbf{b}}(\delta_t; 0)}{s_0(\delta_t; 0)} + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_{\mathbf{b}}(\delta_t; 0)}{s_0(\delta_t; 0)} \right) \sigma^k m_k, \tag{44}$$

where $m_k = \mathbb{E}[v_i^k]$ is known.[60] Note that $\frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_{\mathbf{b}}(\delta_t;0)}{s_0(\delta_t;0)} \right)$ is a known function of $(\delta_{tj})_{j\in\mathbf{J}}$, $(\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$, and $(\lambda_j)_{j\in\mathbf{J}}$. Recall that for $j = 1, ..., J$,

$$\pi_{tj} = \left[\sum_{\mathbf{b}' \in \mathbf{N}_j} \left(\omega_{\mathbf{b}'j} \exp(\delta_{t\mathbf{b}'}) \right)^{1/\lambda_j} \right]^{\lambda_j}. \tag{45}$$

Then, given $(\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$ and suitable regularity conditions, one can uniquely (and locally) back out $(\delta_{tj})_{j=1}^J$ from $(\pi_{tj})_{j=1}^J$:[61]

$$\delta_{tj} = \delta_j(\pi_{t1}, ..., \pi_{tJ}; (\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}). \tag{46}$$

Then, $\frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_{\mathbf{b}}(\delta_t;0)}{s_0(\delta_t;0)} \right)$ is a known function of $(\pi_{tj})_{j\in\mathbf{J}}$, $(\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$, and $(\lambda_j)_{j\in\mathbf{J}}$ and we can recover $\ln \frac{s_{\mathbf{b}}(\delta_t;0)}{s_0(\delta_t;0)}$ using (44) for given σ:

$$\ln \frac{s_{\mathbf{b}}(\delta_t; 0)}{s_0(\delta_t; 0)} = \ln \frac{\mathscr{s}_{t\mathbf{b}}}{\mathscr{s}_{t0}} - \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_{\mathbf{b}}(\delta_t; 0)}{s_0(\delta_t; 0)} \right) \sigma^k m_k. \tag{47}$$

Moreover, (44) implies that for $j = 1, ..., J$,

$$\ln \frac{\mathscr{s}_{tj}}{\mathscr{s}_{t0}} = \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j) \ln \left(\frac{s_j(\delta_t;0)/s_0(\delta_t;0)}{\pi_{tj}} \right) + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_j(\delta_t;0)}{s_0(\delta_t;0)} \right) \sigma^k m_k + \xi_{tj}. \tag{48}$$

In addition to the same linear terms in the C2SLS, the random coefficient also implies non-linear term $\sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_j(\delta_t;0)}{s_0(\delta_t;0)} \right) \sigma^k m_k$ in (48) where σ is unknown. Then, given $\frac{s_j(\delta_t;0)}{s_0(\delta_t;0)}$ and π_{tj} for $j = 1, ..., J$ and all t, $(\delta_1, ..., \delta_J, \alpha, \beta, \lambda_1, ..., \lambda_J)$ and σ can be pinned down by applying the generalized method of moments to (48). As in the standard C2SLS, the resulting solutions should satisfy the second and third equations in (43). Additionally, they should also satisfy (47):

[60] The equality in (44) can be obtained by relying on the real analyticity of PONL model (7) with respect to δ_t around $\delta_t = 0$ and assuming that ν_i has bounded support (or unbounded support with thin tails). Similar real analyticity arguments are derived for mixed logit and probit models in Iaria and Wang (2022).

[61] In practice, the most essential regularity condition is that $\frac{\partial(\pi_{tj})_{j=1}^J}{\partial(\delta_{tj})_{j=1}^J}$ is of full rank at the true values of $(\delta_{tj})_{j=1}^J$, given the true values of $(\Gamma_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$ and of $(\lambda_j)_{j\in\mathbf{J}}$.

$(\delta_1, ..., \delta_J, \alpha, \beta, \lambda_1, ..., \lambda_J, \sigma)$ are such that

$$
\begin{aligned}
\ln \frac{\mathcal{s}_{tj}}{\mathcal{s}_{t0}} &= \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j) \ln \left(\frac{s_j(\delta_t; 0)/s_0(\delta_t; 0)}{\pi_{tj}} \right) + \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_j(\delta_t; 0)}{s_0(\delta_t; 0)} \right) \sigma^k m_k + \xi_{tj}, \\
\ln \frac{s_{\mathbf{b}}(\delta_t; 0)}{s_0(\delta_t; 0)} &= \ln \frac{\mathcal{s}_{t\mathbf{b}}}{\mathcal{s}_{t0}} - \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_{\mathbf{b}}(\delta_t; 0)}{s_0(\delta_t; 0)} \right) \sigma^k m_k, \ \mathbf{b} \in \mathbf{C}_1, \\
\Gamma_t &= (\Gamma_{\mathbf{b}}(\Gamma_{t\mathbf{b}}; \lambda, \pi_t, s(\delta_t; 0)/s_0(\delta_t; 0)))_{\mathbf{b} \in \mathbf{C}_2} \text{ from (16)} \\
\pi_t &= (\phi_j(\pi_t; \lambda, \Gamma_t, s(\delta_t; 0)/s_0(\delta_t; 0)))_{j \in \mathbf{J}} \text{ from (17).}
\end{aligned}
\tag{49}
$$

To identify $(\delta_1, ..., \delta_J, \alpha, \beta, \lambda_1, ..., \lambda_J, \sigma)$ on the basis of (49), one can use similar arguments to those used for the standard PONL model. In estimation, one needs to replace the infinity in the sum of the non-linear term by a finite number K. To guarantee the desired asymptotic properties (at least consistency), similar to the sieve approach, K must increase in T (the number of markets) asymptotically.

Before proposing a convenient iterative estimation procedure based on (49), we illustrate how, similar to the standard PONL model, excluded cost shifters and exogenous product characteristics are appropriate instruments in the presence of random coefficients, particularly for the identification of σ. In practice, we suggest the use of higher-order terms of such instruments (e.g., their polynomials) that can provide additional power with respect to their linear combinations.

Example. As argued above, the last three equations in (49) imply that π_{tj} (and therefore $\ln \pi_{tj}$) is a function of λ and σ: $\ln \pi_{tj} = \ln \pi_j(\lambda, \sigma; \mathcal{s}_t)$. Then, by relying on the first-order Taylor expansion of $\ln \pi_{tj}$ around $\ln \pi_{tj}^0 = \ln \pi_j(\lambda^0, \sigma^0; \mathcal{s}_t)$, and the fact that ν_i is centered ($m_1 = 0$), we can re-write the first equation in (49) as:

$$
\begin{aligned}
\ln \frac{\mathcal{s}_{tj}}{\mathcal{s}_{t0}} &= \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j) \ln s_{(j|j)}(\delta_t; 0) + \sum_{k=2}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_j(\delta_t; 0)}{s_0(\delta_t; 0)} \right) \sigma^k m_k + \xi_{tj} \\
&\quad - (1-\lambda_j) \left[\frac{s_0(\delta_t; 0)}{s_j(\delta_t; 0)} \frac{\partial \pi_j(\lambda^0, \sigma^0; \mathcal{s}_t)}{\partial \lambda} (\lambda - \lambda^0) + \frac{\partial \pi_j(\lambda^0, \sigma^0; \mathcal{s}_t)}{\partial \sigma} (\sigma - \sigma^0) \right],
\end{aligned}
\tag{50}
$$

where $(1-\lambda_j) \ln s_{(j|j)}(\delta_t; 0)$ and $\sum_{k=2}^{\infty} \frac{1}{k!} \frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_j(\delta_t;0)}{s_0(\delta_t;0)} \right) \sigma^k m_k$ are the leading terms. Similar to the standard PONL model, the identification of λ_j can be achieved by exogenous variables (e.g., cost shifters, exogenous product characteristics) that shift $\ln s_{(j|j)}(\delta_t; 0)$. Analogously, to identify σ, one needs instruments that shift $\frac{\partial^k}{(\partial\sigma)^k} \left(\ln \frac{s_j(\delta_t;0)}{s_0(\delta_t;0)} \right)$. Routinely used instruments are valid for this purpose. Let's take $k = 2$ for example: using $m_1 = 0$ and therefore $\frac{\partial s_j(\delta_t;0)}{\partial \sigma} = 0$, we obtain:

$$
\frac{\partial^2}{(\partial\sigma)^2} \left(\ln \frac{s_{tj}(0)}{s_{t0}(0)} \right) = \frac{\partial^2 \ln s_{tj}(0)}{(\partial\sigma)^2} - \frac{\partial^2 \ln s_{t0}(0)}{(\partial\sigma)^2} = m_2 \sum_{\mathbf{b}, \mathbf{b}' \in \mathbf{C}_1} p_{t\mathbf{b}} p_{t\mathbf{b}'} \left(\frac{1}{s_{tj}(0)} \frac{\partial^2 s_{tj}(0)}{\partial \delta_{t\mathbf{b}} \partial \delta_{t\mathbf{b}'}} - \frac{1}{s_{t0}(0)} \frac{\partial^2 s_{t0}(0)}{\partial \delta_{t\mathbf{b}} \partial \delta_{t\mathbf{b}'}} \right).
$$

In general, $\frac{\partial^2}{(\partial\sigma)^2}\left(\ln\frac{s_j(\delta_t;0)}{s_0(\delta_t;0)}\right)$ is not a constant function of p_{tj} (which enters through $p_{t\mathbf{b}}p_{t\mathbf{b}'}$) and x_{tj} (which enters through $\frac{1}{s_j(\delta_t;0)}\frac{\partial^2 s_j(\delta_t;0)}{\partial\delta_{t\mathbf{b}}\partial\delta_{t\mathbf{b}'}} - \frac{1}{s_0(\delta_t;0)}\frac{\partial^2 s_0(\delta_t;0)}{\partial\delta_{t\mathbf{b}}\partial\delta_{t\mathbf{b}'}}$ when $j \in \mathbf{b}$ or $j \in \mathbf{b}'$). As a result, excluded cost shifters for product j and exogenous product characteristics can shift $\frac{\partial^2}{(\partial\sigma)^2}\left(\ln\frac{s_j(\delta_t;0)}{s_0(\delta_t;0)}\right)$, providing identification power for σ.

Remark 1. *Salanié and Wolak (2019) propose a similar approach to ours and rely on a second-order Taylor expansion to approximately estimate σ^2 in a classic BLP model with random coefficients. While both methods can be used to investigate whether σ is close to 0 (as we do in our empirical application),*[62] *our approach is theoretically exact—rather than approximative—thanks to the real analytic property of the PONL model (see footnote 60). Differently, Salanié and Wolak (2019) work with second-order Taylor expansions of the inverse demand functions (eq. 5, page 13), whose real analytic properties with respect to σ are harder to establish.*

Remark 2 (Multiple random coefficients). *In this case, the expansion in (49) can be derived with respect to each σ_ℓ with $\ell = 1, ..., L$, and their cross terms, where L is the number of random coefficients $\{\nu_{i\ell}\}_{\ell=1}^{L}$ and σ_ℓ is the parameter corresponding to $\nu_{i\ell}$. If one, as commonly done in applied work, assumes independence among the random coefficients $\{\nu_{i\ell}\}_{\ell=1}^{L}$, then the Taylor expansion will not include cross terms, greatly simplifying estimation.*

Remark 3 (Nonparametric F). *When F is unknown and to be identified/estimated, we can normalize $\sigma = 1$. One can treat each moment m_k in (49) as a parameter and identify F from its moments.*[63] *Interestingly, in such case the first equation in (49) becomes linear in m_k and estimation simplifies.*

Formulation (49) implies an iterative estimation procedure very similar to that for the standard PONL model without any random coefficient. To illustrate this, consider only the first two terms of the Taylor expansion. Because $m_1 = 0$ and $m_2 = 1$, we then obtain the following iterative estimation procedure:

Step 1. Set $r = 0$ and $(\sigma^2)^{(r)} = 0$. Initialize $\left(\alpha^{(r)}, \beta^{(r)}, \delta_1^{(r)}, ..., \delta_J^{(r)}, \lambda_1^{(r)}, ..., \lambda_J^{(r)}\right)$ and $\left(\Gamma_{t\mathbf{b}}^{(r)}\right)_{\mathbf{b},t}$ to be the solutions of the PONL model without random coefficients and compute $\left(\pi_{tj}^{(r)}\right)_{t,j}$.

Step 2. Compute $h_{t\mathbf{b}}^{(r)} = \frac{1}{2}\frac{\partial}{(\partial\sigma)^2}\left(\ln\frac{s_{\mathbf{b}}(\delta_t;0)}{s_0(\delta_t;0)}\right)$ using $(\lambda_1^{(r)}, ..., \lambda_J^{(r)})$, $\left(\Gamma_{t\mathbf{b}}^{(r)}\right)_{\mathbf{b},t}$, and $(\pi_{tj}^{(r)})_{t,j}$. Compute $\ln\left(\frac{s_{\mathbf{b}}^{(r)}(\delta_t;0)}{s_0^{(r)}(\delta_t;0)}\right) = \ln\frac{\mathfrak{s}_{t\mathbf{b}}}{\mathfrak{s}_{t0}} - h_{t\mathbf{b}}^{(r)} \times (\sigma^2)^{(r)}$.

[62] See also Chesher and Santos Silva (2002), who also use a second-order Taylor expansion to test for the importance of taste variation in a mixed logit model (Heterogeneity Adjusted Logit).

[63] To identify F from its moments (i.e., determinacy of the moment problem), one can for example rely on Carleman's condition.

Step 3. Obtain $\left(\alpha^{(r+1)}, \beta^{(r+1)}, \delta_1^{(r+1)}, ..., \delta_J^{(r+1)}, \lambda_1^{(r+1)}, ..., \lambda_J^{(r+1)}, (\sigma^2)^{(r+1)}\right)$ from the following 2SLS regression:

$$\ln \frac{\mathscr{s}_{tj}}{\mathscr{s}_{t0}} = \delta_j + x_{tj}\beta - \alpha p_{tj} + (1-\lambda_j) \ln \left(\frac{s_j^{(r)}(\delta_t; 0)/s_0^{(r)}(\delta_t; 0)}{\pi_{tj}^{(r)}} \right) + h_{tj}^{(r)}\sigma^2 + \xi_{tj}.$$

Step 4. Compute $\left(\Gamma_{t\mathbf{b}}^{(r+1)}\right)_{\mathbf{b},t}$ and $\left(\pi_{tj}^{(r+1)}\right)_{t,j}$ using $\frac{s_{\mathbf{b}}^{(r)}(\delta_t;0)}{s_0^{(r)}(\delta_t;0)}$ and $\left(\lambda_1^{(r+1)}, ..., \lambda_J^{(r+1)}\right)$ from the standard PONL model without random coefficients.

Step 5. Set $r = r + 1$ and go back to Step 2.

F Monte Carlo Simulations

We investigate by simulation the finite sample performance of the C2SLS with respect to choice set size C and number of iterations of iterative procedure $\bar{K}$.

F.1 Data Generating Process

We generate data from a PONL model with $J = 10$ products and bundles that combine multiple units of 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 $j_k \in \{0\} \cup \mathbf{J}$, $k = 1, 2, 3$.[64] We consider choice sets 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 $((\delta_j = 1, \lambda_j = 0.4)_{j=1}^{10}, \beta = 2, \alpha = 2)$, $\Gamma_{t\mathbf{b}} \sim N(0, 0.1)$, $\xi_{tj} \sim N(0, 0.2)$, $\log(x_{tj}) \sim N(1, 0.1)$ (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 $p_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} p_{tj}$. The monopolist faces no technological advantage or disadvantage in selling bundles: the marginal cost of any bundle $\mathbf{b}$ is the sum of the marginal costs of the units it includes, $z_{t\mathbf{b}} = \sum_{j \in \mathbf{b}} z_{tj}$.

[64] 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.

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. We use two types of instruments. 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 $\mathbf{N}_j \setminus \{j\}$.[65]

F.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 4: Median of RMSEs of C2SLS Estimator

(a) Parameters $(\delta, \beta, \alpha, \lambda)$, $T = 200$ markets (b) Choice Set Size $C = 286$

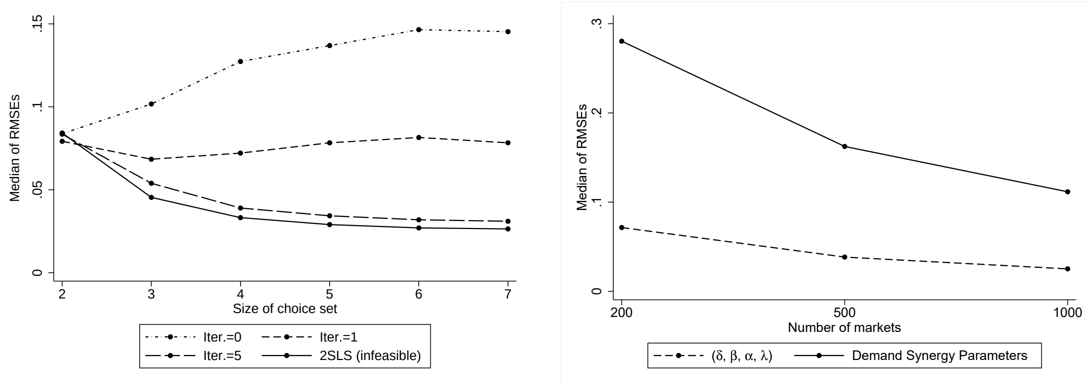


Figure 4 summarizes our simulation results. Figure 4(a) illustrates results for the estimation of $(\delta, \beta, \alpha, \lambda)$ in different scenarios with choice set size $C \in \{66,$ $286, 1001, 3003, 8008, 19448\}$. For each C, we simulate 100 datasets with $T = 200$ markets and then average estimates across these. We summarize the finite sample performance of each estimator in terms of its median Root Mean Square Error

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

(RMSE).[66] 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 details), iteration 1, and iteration 5.

Figure 4(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 4(b) illustrates 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 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.

G Dealing with Products with Undefined $\mathfrak{s}_{tj}$

Without further assumptions, the C2SLS estimator cannot pin down the demand synergy $\Gamma_{t\mathbf{b}}$ of a bundle that includes products not observed to be purchased as single units in market t. This can be seen in (16): if product j' is not observed to be purchased as a single unit in market t, then $\mathfrak{s}_{tj'}$ is not defined and 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 prominent, one can simply exclude them from the analysis. However, when there are many of these bundles, excluding them may correspond to dropping a large share of purchases. In this Appendix, we provide a practical solution to this problem that does not involve excluding bundles from the analysis or any modification of the C2SLS estimator.

The main idea of the proposed approach consists of three steps. In the first step, we "separate away" from bundles any sub-bundle collecting products whose purchase probability of a single unit $\mathfrak{s}_{tj}$ is defined. In the second step, we implement

[66] 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.

the C2SLS estimator only on the products observed to be purchased as single units (i.e., with defined $\mathcal{s}_{tj}$) 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 $\mathcal{s}_{tj}$ is *not* defined.

Suppose the J' products in $\mathbf{J}'$, indexed by $j' = 1, ..., J'$, are only observed to be purchased as part of bundles, but not as single units. These products have undefined $\mathcal{s}_{tj'}$. 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 $(\mathbf{b}_{j'})_{j'=1}^{J'}$.[67] To save on notation, we use the symbol $\mathbf{b}_{-\mathbf{J}'}$ also to refer to the original bundles $\mathbf{b}$ that do not include any purchase of products in $\mathbf{J}'$. 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:

$$\delta_{\mathbf{b}_{j'}} + \xi_{t\mathbf{b}_{j'}} = \ln(\mathcal{s}_{t\mathbf{b}_{j'}}) - \ln(\mathcal{s}_{t0}) - x_{t\mathbf{b}_{j'}}\beta + \alpha p_{t\mathbf{b}_{j'}} - (1-\lambda_j)\ln\left(\mathcal{s}_{t(\mathbf{b}_{j'}|j')}\right), \quad (51)$$

where $\mathcal{s}_{t(\mathbf{b}_{j'}|j')}$ is known because of the 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'.

After having recovered $\delta_{\mathbf{b}_{j'}} + \xi_{t\mathbf{b}_{j'}}$ from (51) 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 H. The results presented in the empirical application in section 5 rely on this procedure. However, in unreported robustness checks, we repeated the empirical analysis by excluding all bundles 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 unchanged.

[67]For each bundle $\mathbf{b}$ and product $j' \in \mathbf{J}'$, $\mathbf{b}_{j'}$ could be empty if $\mathbf{b}$ does not include any unit of j', $\mathbf{b}_{j'} = j'$ if $\mathbf{b}$ includes 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.

H Empirical Application

H.1 Elasticities in Table 4

To simplify exposition, we drop the indexes of household size hs and of market t. Here we derive the demand elasticities we report in Table 4: the percentage changes in the collective number of units purchased in terms of single-unit products in $\mathbf{J}$ and of 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 $\mathbf{A}, \mathbf{B} \in \{\mathbf{J}, \mathbf{C}_2\}$.

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

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

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

H.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 second element $m_{\mathbf{b}2} = 1$ and fifth element $m_{\mathbf{b}(2,2)} = 0$.

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 they 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 sold by the same producer, or 0 otherwise. For example, $o_{1,2} = 1$

if products 1 and 2 are sold by the same producer. Moreover, $o_{1,(1,1)} = 1$ because multiple units of the same product are 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:

$$F(p_{\mathbf{J}_1\cup\mathbf{J}_2}) = \left[\mathbf{O}\circ\left(\mathbf{M}^{\mathrm{T}}\frac{\partial \mathscr{s}_{\mathbf{C}_1}}{\partial p_{\mathbf{C}_1}}\mathbf{M}\right)\right](p_{\mathbf{J}_1\cup\mathbf{J}_2} - c_{\mathbf{J}_1\cup\mathbf{J}_2}) + \mathbf{O}^{\mathrm{T}}\mathscr{s}_{\mathbf{C}_1} = 0_{(|\mathbf{J}_1|+|\mathbf{J}|_2)\times 1}, \quad (52)$$

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 (52) 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 prices 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 $\mathbf{b}$:

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

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 (52):

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

H.3 Counterfactual Simulation: Linear Pricing

To simulate the counterfactual scenario with linear pricing, we start from the setting in Appendix H.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 (25), so that $\tilde{\Gamma}_{\mathbf{b}}^{hs} = \tilde{\gamma}_{\mathbf{b}}^{hs}$ for $hs\in\{\text{single}, \text{multi}\}$ and $\mathbf{b}\in\mathbf{C}_2$, and by letting producers re-optimize with respect to $p_{\mathbf{J}_1} = (p_j)_{j\in\mathbf{J}_1}$.

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:

$$M_{12}^{*\mathrm{T}} F(M_{12} p^*_{\mathbf{J}_1}) = 0_{|\mathbf{J}_1|\times 1}, \tag{53}$$

where $F(\cdot)$ is defined in (52) and $c_{\mathbf{J}_1\cup\mathbf{J}_2}$ is the vector of marginal costs obtained in Appendix H.2 (we assume that marginal costs are not affected by the pricing scheme). We consider any solution to the non-linear system of FOCs (53) 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 one solution to FOCs (53) in each market.

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

In this section, we detail the computation of the entries of Tables 6, 7, and 8. Here we only discuss the computation of absolute changes (in \$, liters (L), or oz), but relative changes (in %) are obtained analogously. 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 4 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:

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

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:

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

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

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

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

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

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:

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

and that of non-sugary CSDs is:

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

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

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

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

Compensating variation. In the setting of PONL model (1), income effects enter linearly into the indirect utilities $U_{it\mathbf{b}}$ for all $\mathbf{b} \in \mathbf{C}$. As a consequence, for households of size $hs \in \{\text{single}, \text{multi}\}$, the compensating variation is:

$$\mathrm{CV}_t^{hs} = \frac{\mathrm{CS}_t^{hs,\text{observed}} - \mathrm{CS}_t^{hs,\text{linear}}}{\alpha^{hs}}, \tag{54}$$

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

$$\mathrm{CS}_t^{hs,d} = \ln\left(\sum_{\ell=0}^{J}\left(\sum_{\mathbf{b}'\in\mathbf{N}_\ell}(\omega_{\mathbf{b}'\ell}\exp(\delta_{t\mathbf{b}'}^{hs,d}))^{1/\lambda_\ell^{hs}}\right)^{\lambda_\ell^{hs}}\right) \tag{55}$$

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

$$\mathrm{CV}_t = \frac{\sum_{hs} w_{hs}\alpha^{hs}\mathrm{CV}_t^{hs}}{\sum_{hs} w_{hs}\alpha^{hs}}. \tag{56}$$

Finally,

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

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 $(\delta_{t\mathbf{b}}^{hs,\text{observed}})_{\mathbf{b}\in\mathbf{C}_2}$ is unchanged. Then, for households of size hs, the compensating variation due to the changes in $(p_{tj})_{j\in\mathbf{J}}$ is:

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

while that due to the changes in $(p_{t\mathbf{b}})_{\mathbf{b}\in\mathbf{C}_2}$ is $\mathrm{CV}_t^{hs} - \mathrm{CV}_t^{hs,\text{single unit}}$. Their average across household sizes is then defined as in (56).

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

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

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

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

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

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

Simulation of sugar tax. Suppose that the sugar tax is t dollars per ounce of added sugar. Then, the sugar tax for n units (liters) of CSD j with τ_j ounces of added sugar is $t\tau_j n$. Denote by $\mathbf{t}$ the vector of sugar taxes corresponding to the single and multiple units of all products in $\mathbf{J}_1 \cup \mathbf{J}_2$. Then, similar to (52), the FOCs in the presence of the sugar tax are:

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

where $p^*_{\mathbf{J}_1\cup\mathbf{J}_2}$ are the new equilibrium prices (net of the sugar tax) to solve for.

H.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 (55), 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 (54), the compensating variation of a change in prices from $(p_1, p_{(1,1)})$ to $(p_1 + \Delta_1, p_{(1,1)} + \Delta_{(1,1)})$ is 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 2, in the observed scenario with quantity discounts, multi-person households are far more likely than single-person households to purchase multiple units of CSDs, so that $\jmath_{(1,1)}^{\text{multi}} > \jmath_{(1,1)}^{\text{single}}$ and $\jmath_1^{\text{multi}} < \jmath_1^{\text{single}}$. 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 clarifies 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 $\jmath_1^{\text{single}}$ single-person households place on $\Delta_1 < 0$ and the larger weight $\jmath_{(1,1)}^{\text{multi}}$ multi-person households instead place on $\Delta_{(1,1)} > 0$.

H.6 Additional Tables

Table 9: Demand Estimates: Demand Synergy Parameters

	hs = single	hs = multi
Average (std. dev.) $\tilde{\gamma}_{t\mathbf{b}}^{hs}$	36.457	56.723
	(39.595)	(55.926)
OLS Estimation Results		
# different producers in $\mathbf{b} = 2$	−2.426	−1.131
	(0.526)	(0.109)
# different producers in $\mathbf{b} = 3$	−4.778	−1.286
	(2.135)	(0.299)
# different producers in $\mathbf{b} = 4$	−8.127	−4.311
	(9.823)	(1.575)
# different products in $\mathbf{b} = 2$	1.969	−0.484
	(0.426)	(0.098)
# different products in $\mathbf{b} = 3$	6.690	1.291
	(1.151)	(0.166)
# different products in $\mathbf{b} = 4$	7.914	2.815
	(3.375)	(0.320)
# different products in $\mathbf{b} = 5$	8.607	6.804
	(9.476)	(0.665)
# different products in $\mathbf{b} = 6$	−	5.584
		(1.085)
# different products in $\mathbf{b} = 7$	−	12.845
		(1.850)
# different products in $\mathbf{b} = 8$	−	20.536
		(3.785)
# different products in $\mathbf{b} = 9$	−	47.184
		(5.378)
$\|\mathbf{b}\|$ fixed effects	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 OLS regression of the demand synergy parameters as obtained from the 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.