The effects of banning loss-leader pricing in grocery retailing markets

[VERY PRELIMINARY. Please do not cite or circulate without permission.]

Jorge Florez-Acosta and Daniel Herrera-Araujo*

This version: February 2022

Abstract

Currently, half of the states of the U.S. ban loss-leader pricing because it is considered predatory. However, recent evidence challenges this view and suggests the ban may rather be inconvenient. We examine the effects of banning loss-leader pricing in grocery retailing markets on supermarket pricing, competition and welfare. To this end, we use scanner data on supermarket sales in the United States and carry out two empirical exercises. First, we estimate the effects of the ban on supermarket chains' prices by exploiting variation on both the policy across states and chains' exposure to the ban across supermarket chains. Our preliminary results suggest that prices significantly increase with exposure to the ban at the chain level, implying that prices at stores located in ban-free states are also higher compared to less exposed chains. Second, we are working on the development and estimation of structural models of demand and supply of multiple products. In this iteration of the paper, we present our demand model and discuss a challenge to identification that is common in the estimation of this type of models. We propose a novel solution to this challenge and show, based on Monte Carlo simulations, that our solution outperforms strategies used by previous literature.

Keywords: loss-leader pricing, below cost pricing, predation, multiproduct retailing, one-stop shopping, multistop shopping, multiproduct demand, shopping costs, demand estimation.

JEL Codes: D12, L13, L81.

Acknowledgments: For helpful discussions, comments and suggestions we thank conference participants at the EARIE Annual Conference (2021), LACEA-LAMES Annual Conference (2021), and Banco de la República Research Network Conference (2021). Renzo Clavijo provided excellent research assistance. This paper uses the Nielsen datasets supplied by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The results in this paper are the researchers' own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. All errors and omissions remain the responsibility of the authors.

*Florez-Acosta: Banco de la República and Universidad del Rosario, Bogotá, Colombia (e-mail: jfloreac@banrep.gov.co); Herrera-Araujo: Université Paris-Dauphine, LEDA (CGEMP), Place du Maréchal de Lattre de Tassigny, (e-mail: daniel.herrera@dauphine.psl.eu).

1 Introduction

Large supermarket chains often price some of their competitive products at or below cost in order to attract consumers into the store and profit from their purchases of products with positive markups. This practice, best known as *loss-leader pricing*, has been banned in a several countries and half of the U.S. states because it raises concerns of predatory conduct among competition authorities. However, anecdotal and economic evidence suggests that the use of loss-leading need not be consistent with predatory conduct, as the main features of predation, namely, losses due to below-cost prices, exclusion of competitors and the subsequent recoup stage are hardly observed in markets where powerful firms use loss-leader pricing (see, for example, Florez-Acosta and Herrera-Araujo (2020)). Recent theoretical literature provides formalization of the alternative view of loss-leader pricing being not predatory and shows that such a practice may even be welfare enhancing (Chen and Rey, 2012, 2019; Johnson, 2017). These results rise new concerns about loss-leading bans as it may lead to higher prices in equilibrium and prevent retail stores from decreasing the price of some of their products below certain threshold during sales not to be accused of charging "too low" prices.

This paper examines the effects of banning loss-leader pricing in grocery retailing markets on supermarket pricing, competition and consumer welfare. To this end, we use scanner data on supermarket sales in the United States and carry out two empirical exercises. First, we estimate the effects of the ban on supermarket chains' prices by exploiting variation on both the policy across states and chains' exposure to the ban across supermarket chains. We complete this exercise for several product categories that are widely used by supermarkets as either permanent or temporary loss-leaders (e.g., milk, beer, tuna and oatmeal). Our preliminary results suggest that prices significantly increase with exposure to the ban at the chain level, implying that prices at stores located in ban-free states are also higher compared to less exposed chains. Second, we estimate structural models of demand and supply of multiple products. On the demand side, we build on (Florez-Acosta and Herrera-Araujo, 2020) and estimate a multiple-discrete choice model in the context of competition between supermarkets that offer the same product line to the same customers. Consumers can purchase bundles of products from either a single store (*one-stop shopping*) or multiple stores (*multistop shopping*) during a given period and incur shopping costs. On the supply side, we are currently working on a model of competition between supermarket chains in a context of uniform pricing (see, DellaVigna and Gentzkow (2019)); that is, stores that are owned by multi-state supermarket chains may not adjust their prices in response to the local market conditions but rather as a function of a chain-wide objective function.[1]

Our general empirical strategy combines two main features: on the supply side, we will focus on states of the U.S. that meet two criteria: first, loss-leader pricing has not been banned, and second, all of the chains present in those states are only present in no-ban states (to avoid the spill-over effects of the ban due to a chain's exposure to ban states). This will allow us to model

[1] We will introduce the supply side in the next iteration of our paper.

the supply side as if supermarkets were totally free to set any price for their products. This implies that the price of some products may lie below cost in equilibrium. On the demand side, our strategy is to estimate a flexible model that allows for the choice of multiple products from multiple stores using standard techniques from the discrete-choice literature. We specify the utility of each product as a function of observed and unobserved product and store characteristics, as well as parameters to be estimated. On every shopping occasion, each consumer faces idiosyncratic shopping costs that increase with the number of supermarkets visited. In line with previous literature, we define *shopping costs* as all of the consumer's real or perceived costs of using a supplier. These may include transportation costs and opportunity costs related to time spent parking, selecting products in the store, and waiting in line at the checkout; they may also account for the taste for shopping (Klemperer, 1992; Chen and Rey, 2012, 2019; Florez-Acosta and Herrera-Araujo, 2020). Each consumer weighs up the extra benefits of dealing with an additional store against the additional costs involved. If the benefits exceed the costs, the individual will visit an additional supermarket. Otherwise, she will make all her purchases at a single location. The total utility of a basket of products is the sum of the product-specific utilities minus the shopping costs.

A common challenge to the estimation of demand for multiple products is the large size of the set of products and stores available to consumers. Previous literature has dealt with this shortcoming by following one of two approaches: first, reducing the set of alternatives by arbitrarily selecting a subset of products (i.e., the "included products") and leaving both the remaining observed (i.e., the "excluded") and the unobserved products as part of the outside good (e.g., Florez-Acosta and Herrera-Araujo (2020)). Second, by aggregating up individual products to the category or the macro category levels (e.g., Thomassen et al. (2017)). Either approach may lead to bias in the demand estimates.

In this paper, we are interested in modeling the demand for multiple individual products (i.e., the former approach). In such a context, the source of the bias is as follows. When excluding an arbitrary number of observed products, the resulting choice set (i.e., the set of all exclusive and exhaustive alternatives) will consist of all of the possible combinations of included products and stores only (and, probably, the outside good). However, this rules out the fact that one included product can be purchased jointly with many other excluded products that are observed in the data. Therefore, a choice set that is composed exclusively by combinations of included products will not actually be exhaustive; as a consequence, the probability that a consumer chooses a given product will be underestimated. Specially, this bias may affect the shopping costs estimates. In fact, given that the shopping costs help rationalize the bundle composition (as it captures complementarities between *a priori* unrelated products-stores), excluding the bundles consisting of combinations of included and excluded implies that the researcher is not fuly accounting for the cross-products and cross-stores complementarities generated by the opportunity costs of shopping.

We develop a novel approach to deal with this shortcoming that allows us to work with a reduced set of individual products and stores from the consumers' full choice set and correct the bias that arises from excluding observed products and stores. We focus on the total probability of purchasing a product, which is a function of the utility of both included and excluded products. Next, for each individual consumer observed in our data, we exploit their full purchase history for a number of periods before a given purchase to compute two empirical probabilities: the probability of choosing any excluded products conditional on choosing any included products; and, on the other hand, the probability of choosing any set of excluded products alone. We plug these empirical probabilities into our total probability of choosing a given bundle of included products, which nonparametrically accounts for the potential complementarities between included and excluded products and, hence, correct the bias introduced by the restricted choice set. Furthermore, the number of parameters to be estimated remains the same.

We test our approach by performing Monte Carlo simulations. We consider a setting with three supermarkets each of which sells the same three products. We allow consumers to purchase any combination of products-stores or to opt for the no purchase option. Our results show that the uncorrected approach leads, in effect, to biased estimates; where the bigger bias is on the coefficients related to the shopping costs. Once we introduce our bias correction, our results show that our estimates consistently estimated.

This paper is structured as follows: the second section offers a review of the literature related to our work. Next, we present an overview of our data and perform a set of reduced-form analyses on the effects of banning loss-leading on average prices at the store and chain levels. Further, in section four we present our model of demand give details on our empirical implementation and estimation and show simulation results from Monte Carlo experiments. Finally, section five concludes.

2 Related literature

This paper relates to several strands of literature. First, a literature, mainly theoretical, that provides explanations for why multiproduct firms charge low prices on some products and high prices on other products (Holton, 1957; Gerstner and Hess, 1990; Lal and Matutes, 1994; Simester, 1995; Lazear, 1995; Verboven, 1999; Ellison, 2005; Gabaix and Laibson, 2006). Second, a strand of literature that studies the economic effects of loss-leader pricing from a theoretical perspective. Hess and Gerstner (1987) examine loss-leader pricing combined with *rain checks* and evaluate the suitability of banning stock outs of advertised sale products. Chen and Rey (2012) study competition between large multiproduct retailers and specicialized stores in a context of heterogeneous consumers that incur shopping costs; they show that loss leading is exploitative rather than predatory. Johnson (2017) also shows that loss-leading can be nonpredatoy using a setting in which large and small multiproduct retailers compete for one-stop shoppers that are uncertain about which products they will buy in future visits to stores. Finally, Chen and Rey (2019) study competition between equally sized multiproduct retailers in a context of heterogeneous consumers that incur shopping costs and show that loss-leading strategies and cross-subsidies are not predatory, and the latter might even be welfare enhancing.

This paper also relates to the literature that provides empirical evidence on the existence of loss-leader pricing on a number of retail markets such as the new car market (Verboven, 1999), the grocery retailing markets (Walters, 1988; Walters and MacKenzie, 1988; Chevalier, Kashyap and Rossi, 2003), and the internet commerce (Ellison and Ellison, 2009). Moreover, this paper relates to a literature that studies seasonal patterns of prices and price promotions in the U.S. retailing sector (Warner and Barsky, 1995; Johansen, 2000; Chevalier, Kashyap and Rossi, 2003; Gagnon and López-Salido, 2020) and a more recent literature that documents the existence of uniform pricing in the U.S. retailing chain sector (DellaVigna and Gentzkow, 2019; Hitsch, Hortaçsu and Lin, 2019), and the transmission of shocks (García-Lebergman, 2021).

Furthermore, this paper closely relates to a strand of literature investigating the effects of below-cost pricing laws on gasoline prices. Anderson and Johnson (1999) uses weekly data on gasoline prices from 42 major cities of the U.S. from March 1992 through December 1993, and finds that average gasoline prices and retail margins were higher in cities in which below-cost pricing restrictions apply. Skidmore, Peltier and Alm (2005) use monthly data from the 50 states of the U.S. from 1983 through 2002 and find the opposite result: average gasoline prices are lower in states with below-cost pricing bans.[2] Last, Peltier, Skidmore and Milne (2013) extend the latter work by using monthly data on retail and wholesale gasoline prices. They find that below-cost pricing laws stimulate the entry of more competitors and are associated with a decrease in both average wholesale and retail prices. Our paper differs from this literature mainly in two ways: first, our focus is the retailing of groceries rather than gasoline; and second, our modeling and empirical strategies.

On the methodological dimension, this paper closely relates to the literature that estimates demand for bundles of products. Hendel (1999) develops a multiple-discrete choice model to explain how firms choose multiple alternative brands of personal computers. Further, Dubé (2004) applies Hendel's model to the case of carbonated soft drinks. Gentzkow (2007) develops a framework in which similar products can be either substitutes or complements. Wildenbeest (2011) sets out a search cost model in which consumers want to purchase a basket of products in a single stop and care about the total price of the basket. This literature has grown rapidly in recent years. Thomassen et al. (2017), Florez-Acosta and Herrera-Araujo (2020) and Leung and Li (2021) develop demand models of multiproduct and multistore choice with shopping costs, which endogenously capture complementarities between categories/products and helps explain the composition of bundles.[3] Iaria and Wang (2020), Ershov et al. (2021), and Wang (2021)

[2] Their intuition for this result is that such laws appear to favor the entry and permanence of smaller gasoline suppliers as opposed to states without any restrictions in which the market tend to be concentrated by large, vertically integrated firms.

[3] Thomassen et al. (2017) focus on competition between specialized stores and supermarkets and develop a model of demand in which consumers make discrete-continuous choices over multiple *macro-categories* of groceries from up to two stores in each period conditional on their idiosyncratic shopping costs. Similarly, Leung and Li (2021) develop a model of demand for multiple categories of products (including both groceries and non-groceries) and stores and allow for continuous choice of quantity and shopping trip costs. Alternatively, Florez-Acosta and Herrera-Araujo (2020) focus on the role of shopping costs in predicting consumer substitution and shopping patterns in a context of competition between supermarkets of similar size and product range; in their setting, the number of stores visited by a consumer is endogenously determined by a stopping rule involving the extra utility and extra costs involved in visiting an additional store.

build on Gentzkow (2007) and develop frameworks to estimate demand for multiple products with complementarity, each of them focuses on particular challenges that emerge in this context.[4]

Finally, this paper relates to a growing body of empirical literature that models consumer choice problems explicitly accounting for consumer frictions, such as, search (e.g. Hortaçsu and Syverson (2004); Hong and Shum (2006); Kim and Bronnenberg (2010); Wildenbeest (2011); De los Santos, Hortaçsu and Wildenbeest (2012); Moraga-Gonzalez, Sandor and Wildenbeest (2013); Koulayiev (2014); Honka (2014); Dubois and Perrone (2015)); switching (e.g., Shy (2002); Viard (2007); Honka (2014)); and shopping costs (Brief (1967); Aguiar and Hurst (2007); Thomassen et al. (2017); Florez-Acosta and Herrera-Araujo (2020); Leung and Li (2021); Dolfen et al. (2022)).

3 Data, background and preliminary evidence

3.1 Overview of the data

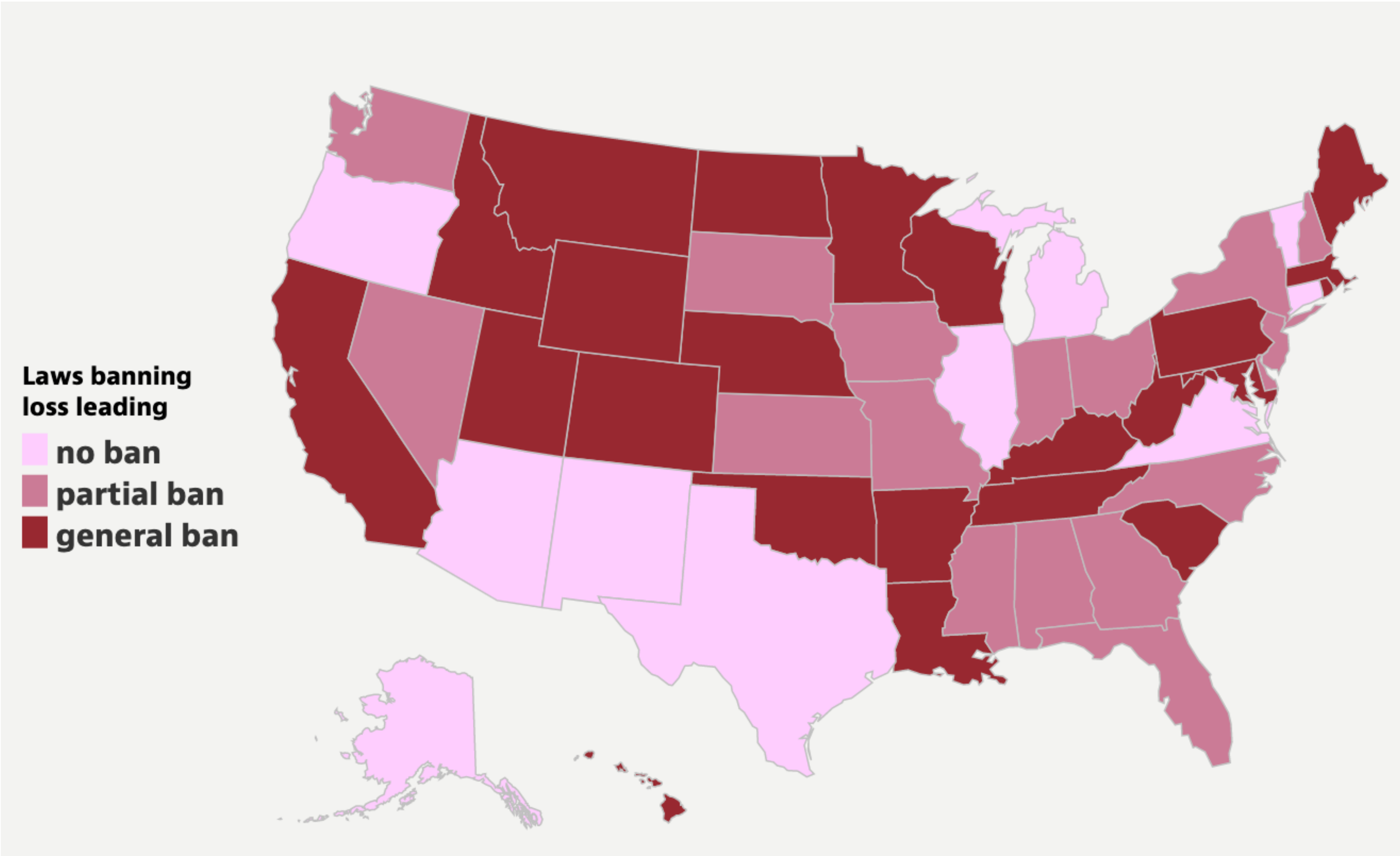
This paper uses the Nielsen's consumer panel and retail scanner data sets supplied by the Kilts Center for Marketing at the University of Chicago Booth School of Business. The consumer panel consists of homescan data that records grocery purchases and household information of a representative sample of households from all states in the United States. On the other hand, the retail data set is scanner data recorded by participating retail stores at the Universal Product Code (UPC) level in a weekly frequency, including retail prices, volume sales, product description, and store location (three first digits of zip code) of every UPC. The two data sets can be merged using both the retailer code and the zip codes.

3.2 Background: loss-leader pricing in the U.S.

Most states of the United States impose restrictions on below-cost pricing either partially on a subset of retail products (e.g., gasoline, cigarettes, milk and alcoholic beverages) or fully on all kinds of retail products. Some of the state laws banning such practice date back from the 1930s. Figure 1 shows the states of the U.S. that currently have a total ban (in dark red), a partial ban (in light red) and no ban (in pink) of loss-leader pricing. States with a total ban include California, Wisconsin, Massachusetts and other 20 states. On the other hand, states with a partial ban include Nevada, New York, Georgia and other 13 states. Table 1 shows the products for which loss-leader pricing has been banned in states in which a partial ban applies. These are mainly cigarettes (banned in 11 states), and gasoline (banned in 6 states). Finally, only 10 states do not impose any restriction of this kind on the pricing of retail products; they include Illinois, Michigan, Texas, Arizona and other 6 states.

[4] Iaria and Wang (2020) and Wang (2021) address the problem of identification of demand models for multiple products when complementarity may emerge; the former provides conditions for identification and an MLE estimator when transaction level (individual) data is available; the latter focuses on aggregate (product-level) data and also provides conditions for identification of the demand parameters. Ershov et al. (2021) deals with large choice sets, which is a common challenge in this literature, and propose an estimator based on aggregate data at the product category level but exploiting micro moments from transaction level data

Figure 1: States of the United States of America in which loss-leader pricing is banned



Notes: the figure shows the States of the US that currently have a total ban (in dark red), a partial ban (in light red) and no ban (in pink) of loss-leader pricing.
Source: Fleisher, Chris, "Loss-leaders: predatory or practical?", American Economic Association, https://www.aeaweb.org/research/loss-leading-bans-retail-competition.

Table 1: Products included in sales-below-cost ban in states with partial restrictions

Included products	Number of states	States
Gasoline	6	Missouri, New York, North Carolina, Georgia, Alabama, Florida.
Cigarettes	11	Washington, Nevada, South Dakota, Iowa, Indianna, Ohio, New York, New Jersey, Delaware, Mississippi.
Milk	3	Nevada (dairy), Missouri, North Carolina.
Alcoholic beverages	2	Kansas, New Hampshire.

Source: Fleisher, Chris, "Loss-leaders: predatory or practical?", American Economic Association, https://www.aeaweb.org/research/loss-leading-bans-retail-competition.

3.3 Preliminary descriptive evidence

In this section, we explore whether the loss-leader ban that is in place in 26 states of the U.S. affects supermarket prices both in the states where the ban applies and in states where the ban does not apply. In effect, there is evidence that most retail chains in the U.S. set uniform prices across stores (DellaVigna and Gentzkow, 2019), which suggests that if the ban policy has any effect on prices, hence, a retail chain with stores in both ban and no ban states should take the ban into account to set their optimal price for a product for all of its stores and to determine the size of a discount for a promotion on a product to avoid being fined for setting too low prices in ban states. In this iteration, we examine the prices of all of the UPCs reported by supermarkets and recorded in the Nielsen's retail scanner data for 2016, for six product categories: milk, lager beer, light beer, ale beer, tuna and oatmeal. These product categories are widely used by supermarkets as loss-leaders (see Chevalier, Kashyap and Rossi (2003)). We carry out these analyses separately by product category.

Table 2 presents summary statistics of some observed characteristics of the supermarket chains in our sample. The number of chains ranges from 74 to 77. In all cases, there is variation both in the number of states where the chain has stores and in the state's ban status–we refer to a state where loss-leader pricing has been banned as "ban state", while to a state where this practice has not been banned as "no ban state". In effect, the share of chains that are present in only one state ranges from 31% to 37%; the share of chains with stores located in ban states only ranges from 21% to 28%; the share of chains with stores located in no ban states only ranges from 26% to 33%; and the share of chains with stores located in both ban and no ban states range from 45% to 47%. Among multi state chains, the number of states where the average chain owns stores ranges from 14 to 23. On average, between 34.04% and 47.66% of a chain's stores are located in ban states; the share of dollar sales in ban states ranges, on average, from 33.6% to 49.3%.

Previous theoretical literature on loss-leader pricing suggests that banning this practice can lead to higher equilibrium prices. Further, the presence of multistate chains that charge uniform

Table 2: Sample composition and summary statistics of the supermarket chains included in our sample

Variable	Milk	Lager beer	Light beer	Ale beer	Tuna	Oatmeal
Panel A. Chains reporting sales						
Number of chains	76	76	77	76	74	74
Share of chains present in a single state	33%	34%	36%	37%	31%	32%
Share of chains present in ban states only	27.6%	21%	22%	21%	24%	26%
Share of chains present in no-ban states only	26.3%	33%	32%	33%	28%	28%
Share of chains present in both ban and no-ban states	46%	46%	45%	46%	47%	46%
Panel B. No. of states where multistate chains are present (excluding sigle state chains)						
Mean	23	17	19	14	19	19
Min	2	2	2	2	2	2
Median	16	15	15	15	16	16
Max	49	49	42	40	49	49
Panel C. Sales revenue (million dollars)						
Mean	3.35	2.30	2.76	1.14	0.48	0.52
p25	1.39	0.74	0.67	0.50	0.15	0.15
Median	2.09	1.90	2.40	0.77	0.24	0.25
p75	3.77	3.73	3.88	1.30	0.49	0.72
Panel D. Share of dollar sales in ban states						
Mean	49.3%	35.7%	33.6%	37.6%	49.0%	47.5%
p25	34.1%	14.2%	16.5%	14.1%	27.1%	25.4%
Median	49.7%	23.1%	25.2%	25.6%	49.2%	48.4%
p75	63.6%	54.1%	47.0%	56.1%	65.9%	64.9%

Table 2: Continued

Variable	Milk	Lager beer	Light beer	Ale beer	Tuna	Oatmeal
Panel E. No. of stores per chain						
Mean	2126	1398	1603	890	1493	1406
Min	1	1	1	1	1	1
Median	868	793	956	466	868	868
Max	8300	7034	5931	5914	8438	8438
Panel F. Share of stores in ban states (all chains)						
Mean	47.64%	35.93%	34.04%	36.63%	47.66%	46.32%
Min	0%	0%	0%	0%	0%	0%
Median	48.24%	31.53%	25.86%	26.01%	48.24%	48.24%
Max	100%	100%	100%	100%	100%	100%
Panel G. Share of stores in ban states (chains present in both ban and no ban states only)						
Mean	48.21%	36.35%	34.21%	36.91%	48.02%	47.01%
Min	1.79%	1.79%	1.79%	1.79%	1.79%	1.79%
Median	48.24%	31.53%	26.01%	36.91%	48.24%	48.24%
Max	98%	98%	98%	98%	98%	98%

Notes: The table shows summary statistics of the supermarket chains included in the Nielsen database for some product categories that we selected. The database records positive sales from stores belonging to supermarket chains.
Source: Nielsen data. Authors' calculations.

prices across stores suggests that stores that are owned by chains that are present both in ban and in no ban states and that are located in no ban states may be indirectly affected by the ban. To test this, we perform two empirical exercises. First, we regress category price indices at the chain level on a measure of the exposure that a retail chain has to the ban policy based on the share of volume sales of that chain in ban states relative to the their total volume sales in all of the states in which that chain is present. And second, we regress a store-level price index on the exposure measure for a sample of stores that are owned by chains present in both ban and no ban states but that are located in no-ban states only (i.e., we exclude stores located in ban states).

We follow Chevalier, Kashyap and Rossi (2003) and compute a store-level category price index as the weighted average of the log of the price of all of the UPCs u, of a category c, sold by store l, located in county, k, of state s, and owned by retail chain r, at week t. This index is given by:

$$P_{clrkst} = \sum_{u \in c} w_{uclrkst} \ln P_{uclrkst}, \tag{1}$$

where $P_{uclrkst}$ is the price of UPC u and $w_{uclrkst}$ is the dollar share of UPC u, in category c, sold by store l, of retail chain r, located in county k, of state s, in week t.

Next, we compute a chain-level category price index as:

$$P_{crkst} = \sum_{l \in r} \sum_{u \in c} w_{uclrkst} \ln P_{uclrkst}, \tag{2}$$

We compute our exposure measure as the sum of volume sales of all products in a specific category sold by all of the stores of a retail chain in all of the ban states where that chain is present divided by the total volume sales of products in that category of that chain's stores across all of the states where they are present. That is:

$$Exposure_{crt} = \frac{\sum_{u,l,k,s} \mathbb{1}_{\{ban\, state\}}(q_{uclrkst})}{\sum_{u,l,k,s} q_{uclrkst}}, \tag{3}$$

where $q_{uclrkst}$ is the total quantity of UPC u, of a category c, sold by store l, of retail chain r, located in county k, of state s, in week t; and $\mathbb{1}_{\{ban\, state\}}$ is an indicator function that takes on 1 if state s bans loss-leader pricing and zero otherwise.

We first explore the potential effects of the ban on prices for our full sample of retail chains (chains present in ban states only, chains present in no-ban states only, and chains present in both ban and no-ban states). To do that, we regress our chain-level price index on our exposure measure and a full set of fixed effects that control for time-varying observed and unobserved chain characteristics, local market (county) structure and demographics, and state level business cycles. Our specification is:

$$P_{crkst} = \beta_0 + \beta_1 Exposure_{crt} + \varsigma_{rt} + \gamma_k + \eta_{st} + \varepsilon_{crkst}, \tag{4}$$

where ς_{rt} is chain-time fixed effects, γ_k is county fixed effects, η_{st} is state-time fixed effects, and

ε_{crkst} is a random shock.

Panel A of Table 4 displays estimates from separate regressions for our six product categories. Results show that for milk and the three categories of beer, the average price of the category tends to be higher the more exposed is the retail chain to states that ban loss-leader pricing. This effect is not significant for tuna and oatmeal which are used as loss leaders during very specific seasons of the year as opposed to milk and beer that are more persistently used as loss-leaders throughout the year (see, Chevalier, Kashyap and Rossi (2003)).

A potential concern with this exercise may be related to the composition of our samples: 54% of chains, on average, are present either in ban states only (i.e., are fully exposed to the ban policy) or in no ban states only (i.e., are not at all exposed to the ban policy); this may be driving our results as fully exposed chains may charge, on average, higher prices than partially and no exposed chains. In order to check this, we perform the same regression as in (4) but excluding both fully exposed and unexposed chains. Panel B of Table 4 presents the results, which are similar to those obtained with the full sample for each product category. In effect, results indicate a similar positive and statistically significant relationship for the chain-level price index of chains that are present in both ban and no ban states. Overall, our results suggest that chains tend to charge higher prices on products that they more often use as loss-leaders as a potential reaction to the ban policy. This is consistent with conventional wisdom according to which too low prices on some specific categories (e.g., staples) are more likely to rise concerns of unlawful conduct from competition authorities.

Concerning chains owning stores in both ban and no ban states, a natural question to ask is whether their stores located in no ban states are affected by the policy of ban states, given that uniform pricing appears to be a widespread practice among most retail chains in the U.S. (see DellaVigna and Gentzkow (2019)). To check this, we regress a store-level price index computed as in equation (1) on our chain-level exposure measure given by equation (3) for the subsample of stores that are owned by these retail chains and are located in no ban states only; that is, we restrict our attention to stores that are not directly exposed to the ban, but that are indirectly exposed because the owning chain has stores in ban states also. We complete this exercise for each of the categories listed above. Our specification is:

$$P_{clrkst} = \lambda_0 + \lambda_1 Exposure_{crt} + \zeta_{lt} + \varsigma_{rt} + \eta_{st} + \omega_{clrkst}, \tag{5}$$

where ζ_{lt} are store-time fixed effects, ς_{rt} are chain-time fixed effects, η_{st} is state-time fixed effects, and ω_{clrkst} is a random shock.

Table 5 presents the results. Similar to the regressions with the full sample of retail chains, we find that average prices of a category at stores located in no ban states tend to be higher the more exposed the owning chain is to states where a ban is in place. This effect is not statistically significant for lager beer, tuna and oatmeal.

The persistent insignificant results for tuna and oatmeal from our previous regressions suggest that the ban policy may not affect all of the products sold by a supermarket chain. This

Table 4: Results from linear regressions of a category price index at the chain level

Variable	Milk	Lager beer	Light beer	Ale beer	Tuna	Oatmeal
Panel A. *All retail chains*						
Exposure	10.20***	1.53*	1.15*	1.56**	-0.96	0.57
	(3.01)	(0.79)	(0.60)	(0.60)	(1.25)	(0.74)
Constant	0.22	7.91***	8.61***	7.90***	1.40**	3.02***
	(1.36)	(0.25)	(0.19)	(0.21)	(0.56)	(0.33)
Chain × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sq	0.20	0.19	0.19	0.18	0.24	0.17
Obs	446,323	289,742	292,436	258,450	399,381	416,363
Panel B. *Retail chains with stores in both ban and no ban states*						
Exposure	10.18***	1.49*	1.10*	1.48***	-0.96	0.58
	(3.04)	(0.79)	(0.61)	(0.61)	(1.26)	(0.74)
Constant	-0.03	7.67***	8.33***	7.66***	1.27**	2.87***
	(1.37)	(0.24)	(0.19)	(0.21)	(0.57)	(0.32)
Chain × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sq	0.17	0.17	0.17	0.16	0.19	0.15
Obs.	424,463	271,380	274,045	240,224	377,729	394,918

Notes: The table shows results from regressions of a category price index computed at the chain level on a measure of exposure of the retail chain to states with a ban of loss-leader pricing. We compute exposure as the sum of volume sales of all products in a specific category combined in a retail chain in all of the states where that chain is present and a ban is in place on the total volume sales of products in that category of that chain's stores across all of the states where it is present. Columns present results from separate regressions for each product category. Panel A presents regression results for a sample that includes all retail chains (chains present in ban states only, chains present in no-ban states only, and chains present in both ban and no-ban states), whereas Panel B presents regression results for a sample that includes chains present in both ban and no-ban states. Standard errors, given in parentheses, are clustered at the retail chain level.*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Nielsen data. Authors' calculations.

is consistent with conventional wisdom and previous evidence according to which tuna and oatmeal are not often used as loss-leaders. However, supermarkets do use these two particular product categories as loss-leaders during certain times of the year. In fact, Chevalier, Kashyap and Rossi (2003) show that tuna and oatmeal experience price decreases during peak demand periods (we refer to these categories as "seasonal" loss-leaders). In a context of loss-leading ban in some states, a relevant question to ask is whether the prices of products that are used as seasonal loss-leaders decrease in the same proportion, in a lower proportion or do not decrease at all relative to the prices of the same products in no ban states.

To examine this question, we regress our chain-level category price index defined in equation

Table 5: Results from linear regressions of a category price index at the store level

Variable	Milk	Lager beer	Light beer	Ale beer	Tuna	Oatmeal
Exposure	2.70***	0.24	0.25**	0.19**	-0.24	0.64
	(0.50)	(0.18)	(0.12)	(0.09)	(0.44)	(0.39)
Constant	0.07	2.18***	2.34***	2.21***	0.30	0.60***
	(0.20)	(0.05)	(0.03)	(0.03)	(0.18)	(0.16)
Store × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Chain × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
State × Month FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sq	0.965	0.861	0.887	0.802	0.798	0.688
Obs.	874,987	659,398	662,717	567,994	767,941	795,266

Notes: The table shows results from regressions of a category price index computed at the store level on a measure of exposure of the retail chain to states with a ban of loss-leader pricing, using a sample of stores located in no ban states that are owned by chains that have stores in ban states also. We compute exposure as the sum of volume sales of all products in a specific category combined in a retail chain in all of the states where that chain is present and a ban is in place on the total volume sales of products in that category of that chain's stores across all of the states where it is present. Columns present results from separate regressions for each product category. Standard errors, given in parentheses, are clustered at the retail chain level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Nielsen data. Authors' calculations.

(2) on a dummy indicating the season of peak demand and its interaction with dummies indicating if a chain is present in both ban and no ban states (which we term "Ban-Noban chain") or if the chain is present in ban states only (which we term "Ban-only chain"). Our specification is:

$$P_{crkst} = \alpha_0 + \alpha_1 S_t + \alpha_2 S_t \times Chain\_BNB_r + \alpha_3 S_t \times Chain\_B_r + \theta_r + \phi_k + \xi_{st} + \nu_{crkst}, \quad (6)$$

where $S_t = 1$ if peak demand season; $S_t = 0$ otherwise; $Chain\_BNB_r = 1$ if chain present in both ban and no ban states, and $Chain\_BNB_r = 0$ otherwise; $Chain\_B_r = 1$ if chain present in ban states only, and $Chain\_B_r = 0$ otherwise, θ_r are chain fixed effects, ϕ_k are county fixed effects, ξ_{st} are state × time fixed effects; and ν_{crkst} is a random shock.

We complete this exercise separately for tuna and oatmeal. Table 6 presents the results. In the case of tuna, our results show that its price decrease during lent, which is consistent with Chevalier, Kashyap and Rossi (2003); however, chains that are present in both ban and no ban states decrease their price by a lower amount compared to chains with stores in no ban states only. This effect is not statistically significant for chains in ban states only. In the case of oatmeal we find similar results for chains in both ban and no ban states, but find a statistically significant result for chains in ban states only; overall, chains that are exposed to the ban policy tend to barely decrease their oatmeal prices during demand peaks compared with chains in no ban states. This evidence is in line with our previous results and suggest that some exposure to the policy may prevent chains to decrease their prices of certain key products during peak demand seasons potentially to avoid an investigation for selling products bellow cost.

Table 6: Results from linear regressions of a category price index at the chain level

Variable	Tuna	Oatmeal
Lent	-0.70***	—
	(0.20)	—
Lent × Ban-Noban chain (=1 if yes)	0.43***	—
	(0.15)	—
Lent × Ban-only chain (=1 if yes)	0.18	—
	(0.21)	—
Cold	—	-0.22***
	—	(0.07)
Cold × Ban-Noban chain (=1 if yes)	—	0.21***
	—	(0.07)
Cold × Ban-only chain (=1 if yes)	—	0.18**
	—	(0.08)
Constant	1.00***	3.28***
	(0.02)	(0.01)
Chain FE	Yes	Yes
County FE	Yes	Yes
State × Month FE	Yes	Yes
R-sq	0.241	0.165
Obs.	399,381	416,363

Notes: The table shows results from regressions of a category price index computed at the retail chain level on a dummy indicating the season of peak demand and its interaction with dummies indicating if a chain is present in both ban and no ban states (i.e., "Ban-Noban chain") or if the chain is present in ban states only (i.e., "Ban-only chain"). The two columns correspond to separate regressions: the first column uses data for tuna, and the second column uses data for oatmeal. Standard errors, given in parentheses, are clustered at the retail chain level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Source: Nielsen data. Authors' calculations.

4 A structural analysis of the effects of banning loss-leader pricing

4.1 The model

Our demand model is based on Florez-Acosta and Herrera-Araujo (2020). In each choice occasion, consumers can purchase baskets of products from either one or multiple stores, conditional on their idiosyncratic shopping costs.

4.1.1 Consumer choice model

There are I consumers in the market indexed by $i = 1, \dots, I$ with idiosyncratic valuations of K grocery products indexed by $k = 1, \dots, K$. Suppose there are three store chains in the market indexed by $r \in \{A, B, C\}$ that supply the same products to all consumers.[5] We observe

[5] Assuming that all consumers have access to the same product range may appear unreasonable. However, this helps us to reduce dimensionality issues in the estimation of the model. An extension of the model would be to

consumers making choices of products and stores on many time periods, indexed by $t = 1, \ldots, T$. Customer i purchasing product k from store r in period t derives a net utility of $\overline{v}_{ikrt}$, which is a function of the price of the product and other characteristics.[6]

Consumers have unit demand for each product class and can purchase one, two or three products in the same period. Let $\mathcal{B}$ be the set of all exclusive and exhaustive baskets. Baskets with multiple products may be purchased from a single store (*one-stop shopping*) or from multiple stores (*multistop shopping*). A consumer favors multistop shopping if her shopping costs are sufficiently small, otherwise she will optimally make her purchases from a single store.

In the formulation of the model, we focus on the fixed component of the total shopping costs that may account for the consumer's taste for shopping. From now on, we will refer to this fixed cost as "shopping costs" and denote it as s_i. Transport costs, which are an important component of the total cost of shopping, are accounted for by including the distance to stores as an additive term to the utility function of a basket of products (see below). Accordingly, shopping costs are assumed to be independent of store characteristics (e.g., size, facilities, location) and time invariant. Furthermore, we assume that s_i is randomly drawn from a continuous distribution function $G(\cdot)$ and positive density $g(\cdot)$ everywhere. Finally, we assume that consumers are well informed regarding prices and product characteristics. Therefore, consumers do not need to engage in a costly search to gather information about prices and product quality.

Consumer i is supposed to exhibit optimal shopping behavior. This implies that she makes an optimal choice involving two elements: whether to be a one- or multistop shopper, and which stores to visit for each of the products she wants to buy. Roughly speaking, the choice set of consumer i will be restricted by the number of stores she can visit given her shopping costs, so that her choice will consist of selecting the mix of products and stores that maximize the overall value of the desired basket. In line with this, a three-stop shopper who can visit all stores and wants the three products will select the best product–store combination from the alternatives existing in the market within each category. A two-stop shopper will select the mix of two stores maximizing the utility of the desired basket from all possible product–store combinations. Her final basket will consist of the best of the two alternatives in each product category. Finally, a one-stop shopper will choose the store offering the largest overall value for the whole basket of products.

Formally, let D_{ir} denote the distance traveled by consumer i from her household location to store r's location, for all $r \in \{A, B, C\}$, and τ denote a parameter that captures the consumer's valuation of the physical and perceived costs of traveling that distance. We define the utility net of transport costs of a shopper who is able to visit only one of the three stores in the market

relax this assumption and allow for heterogeneous choice sets.

[6]For now, we do not specify a functional form for the product-level utility, as it is not necessary for setting out the model. We will assume a parametric specification at the empirical implementation stage in Section 4.2.

as follows:[7]

$$v_{it}^1 = \max\left\{\sum_{k=1}^{K} \overline{v}_{ikAt} - \tau D_{iA}, \sum_{k=1}^{K} \overline{v}_{ikBt} - \tau D_{iB}, \sum_{k=1}^{K} \overline{v}_{ikCt} - \tau D_{iC}\right\}. \tag{7}$$

Similarly, the net utility of a two-stop shopper is given by:

$$\begin{aligned} v_{it}^2 = \max\Bigg\{ & \sum_{k=1}^{K} \max\{\overline{v}_{ikAt}, \overline{v}_{ikBt}\} - \tau(D_{iA} + D_{iB}), \\ & \sum_{k=1}^{K} \max\{\overline{v}_{ikAt}, \overline{v}_{ikCt}\} - \tau(D_{iA} + D_{iC}), \\ & \sum_{k=1}^{K} \max\{\overline{v}_{ikBt}, \overline{v}_{ikCt}\} - \tau(D_{iB} + D_{iC})\Bigg\}. \end{aligned} \tag{8}$$

Finally, the net utility of a consumer who is able to visit all of the stores is given by:

$$v_{it}^3 = \sum_{k=1}^{K} \max\{\overline{v}_{ikAt}, \overline{v}_{ikBt}, \overline{v}_{ikCt}\} - \sum_{r \in \{A,B,C\}} \tau D_{ir}. \tag{9}$$

Note that the expressions in (7), (8), and (9) are particular cases of a more general utility function, in which—conditional on shopping costs—an n-stop shopper selects the subset of stores that maximizes the overall utility of her desired basket. For a one-stop shopper, these subsets are singletons, for a two-stop shopper they contain two elements, and for a three-stop shopper each subset of stores contains precisely the number of stores in the market, which is why she does not need to maximize over subsets of supermarkets.[8]

To determine the number of stops to be made, consumer i weighs the extra utility of undertaking n-stop shopping with the extra costs involved, taking into account the fact that the total cost of shopping increases with the number of stores visited. Let $\delta_{it}^2 \equiv v_{it}^2 - v_{it}^1$ and $\delta_{it}^3 \equiv v_{it}^3 - v_{it}^2$ be the incremental utilities that consumer i derives from visiting, respectively, two stores rather than one and three stores rather than two. Florez-Acosta and Herrera-Araujo (2020) show that consumer i will choose the mix of stores that maximizes her utility conditional on the extra shopping cost being at most the extra utility obtained from visiting additional stores. Hence, the highest possible shopping costs for any consumer able to undertake multistop shopping at either two or three stores, respectively, in equilibrium are given by the following critical cutoff points:

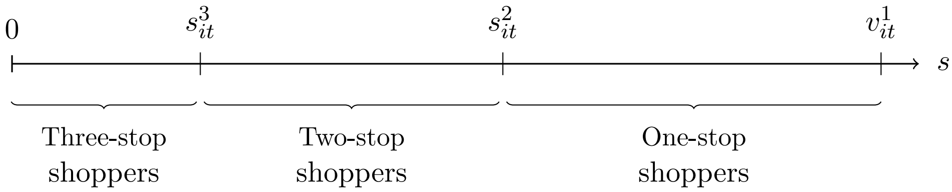
$$\begin{aligned} s_{it}^2 &= \delta_{it}^2, && \text{for two-stop shopping, and} \\ s_{it}^3 &= \delta_{it}^3, && \text{for three-stop shopping.} \end{aligned} \tag{10}$$

[7] Note that the utilities below depend on the vector of all prices of products sold by the three stores in the market, which we denote by $\mathbf{p}_t$. However, we omit this for the sake of simplicity in the presentation of the model.

[8] A general expression of the utility and choices of an n-stop shopper is derived by Florez-Acosta and Herrera-Araujo (2020).

Note that these cutoff points depend on the period of purchase, t. The derived cutoffs for the distribution of shopping costs in (10) indicate that for given shopping costs, consumers only care about the marginal utility of visiting an additional store in making their final decision on how many stores to visit. Moreover, one-, two- and three-stop shopping patterns arise in equilibrium and will be defined over the entire support of $G(\cdot)$ (see Figure 2).

Figure 2: One-, two-, and three-stop shopping



4.1.2 Aggregate demand

Let $\mathcal{B}_2, \mathcal{B}_3 \in \mathcal{B}$ be subsets of baskets involving two- and three-stop shopping, respectively. The aggregate demand for product k supplied by store r is given by:

$$\begin{aligned} q_{krt}(\mathbf{p}_t) = & \left[G(v_{it}^1(\mathbf{p}_t)) - G\left(s_{it}^2(\mathbf{p}_t)\right)\right] P_{irt}^1(\cdot) \\ + & \left[G\left(s_{it}^2(\mathbf{p}_t)\right) - G\left(s_{it}^3(\mathbf{p}_t)\right)\right] \prod_{\{b \in \mathcal{B}_2 \,|\, kr \in b\}} P_{irt}^2(\cdot) \\ & + G\left(s_{it}^3(\mathbf{p}_t)\right) \prod_{\{b \in \mathcal{B}_3 \,|\, kr \in b\}} P_{irt}^3(\cdot), \end{aligned} \tag{11}$$

where $\mathbf{p}_t$ is the $(K * 3) \times 1$ vector of prices of the products sold by the three stores in the market, P_{irt}^1 is the probability that a one-stop shopper decides to shop at store r, P_{irt}^2 is the probability that a two-stop shopper chooses store r as one of the two stores that she will visit, and P_{irt}^3 is the probability that a three-stop shopper decides to select basket b including product kr. All of these probabilities are known by shoppers, and are functions of observable characteristics and parameters. However, for the sake of simplicity we do not specify this dependence at this stage. We defer these details to the empirical section below.

The own- and cross-price elasticities of demand are given by the standard formula $\eta_{krht} = \frac{\partial q_{krt}}{\partial p_{jht}} \frac{p_{jht}}{q_{krt}}$ for all $j \in \{1, \ldots, K\}$, $h \in \{A, B, C\}$. It is important to note that a price change may affect not only the market shares per type of shopper but also the shopping cost cutoff values given that they depend on utilities. As a consequence, the distribution of shoppers between one-, two-, and three-stop shopping groups changes. In fact, an increase in product k's price at store r reduces the indirect utility of consumer i visiting store r. Therefore, she may consider making fewer stops and purchasing a substitute for this product from a rival store, say h, as the gain in utility from visiting an additional store may not be sufficient to offset the extra shopping cost.

4.1.3 Supply

[To be completed.]

4.2 Empirical implementation

4.2.1 Product-level demand

We empirically specify the product-level utility as a function of observed and unobserved product and store characteristics, and time fixed effects. We allow consumer heterogeneity to enter the model through the price coefficient, which is a function of observed and unobserved household characteristics. Formally, let the utility of consumer i from purchasing product k from store r at time t be given by:

$$\overline{v}_{ikrt} = -\alpha_i p_{krt} + \mathbf{x}_{kr}\boldsymbol{\beta} + \xi_{kr} + \phi_t, \tag{12}$$

where p_{krt} is the price of product k at store r, $\mathbf{x}_{kr}$ is a vector of observed product-store characteristics, ξ_{kr} captures the mean valuation of unobserved product-store characteristics, which we capture by including product-store dummy variables, and ϕ_t are time fixed-effects. Finally, α_i is an individual-specific coefficient that captures the marginal utility of price and $\boldsymbol{\beta}$ is a vector of parameters common to all households.

4.2.2 Basket-level demand

In line with our modeling framework, we empirically define the utility that a n-stop shopper ($n \in \{1, 2, 3\}$) derives from purchasing a basket $b \in \mathcal{B}$ as:

$$\begin{aligned} u_{ibt} &= \sum_{kr \in b} \overline{v}_{ikrt} - \tau D_{ir} - s_i n_b + \varepsilon_{ibt}, \\ &= v_{ibt} - s_i n_b + \varepsilon_{ibt}, \end{aligned} \tag{13}$$

where v_{ibt} is the overall utility of basket b net of transport costs as defined by equations (7) through (9) above, n_b is the number of stores visited to purchase basket b, s_i is the individual shopping cost, and ε_{ibt} is an idiosyncratic basket-level shock to utility. We allow price dis-utility, α_i, and shopping costs, s_i, to vary across households, but not across time.

Note that equation (13), along with equations (12), fully specify the utilities of one- and multistop shoppers as a function of price, product characteristics, distance to stores, and individual shopping costs. Thus, our utility accounts for both the vertical and horizontal dimensions of consumers' valuations of products. The vertical component is captured by product-store characteristics, while the horizontal component is captured by distance, which varies across store formats and zip codes, and shopping costs. Finally, we normalize the utility of the "no purchase option" to zero. Thus, it is modeled as a function of an individual random shock to utility, $u_{iOt} = \varepsilon_{iOt}$.

A consumer who wishes to buy a basket b at time t faces a choice set $\mathcal{B}$ of mutually exclusive and exhaustive alternatives consisting of combinations of products and stores. The basket she chooses is such that she obtains the highest possible utility net of shopping costs. This is, for all $b' \in \mathcal{B}$, consumer i chooses basket b at time t if:

$$u_{ibt} > u_{ib't}, \; \forall \, b' \neq b.$$

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \mathbf{s}', \tau)'$ be a vector containing the parameters to be estimated, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)'$ and $\mathbf{s} = (s_1, \ldots, s_I)'$ denote the vector of individual preference parameters. We do not place any restriction on the joint distribution. We follow Dubois, O'Connell and Griffith (2020) and use the large T dimension of our panel to recover estimates of individual specific parameters $(\boldsymbol{\alpha}, \mathbf{s})$, while the large I dimension of our panel allows us to identify nonparametrically the joint probability distribution function $f(\alpha_i, s_i)$ using the empirical probability distribution function of estimated $(\boldsymbol{\alpha}, \mathbf{s})$.

We assume that the random shocks to utility, ε_{ibt}, are distributed i.i.d. type I extreme value. Integrating over ε_{ibt} yields the closed-form choice probability of basket b, at time t, as a function of the characteristics of products and supermarkets:

$$P_{ibt}(\mathbf{X}, \mathbf{p}_t; \boldsymbol{\theta}) = \frac{\exp(v_{ibt} - s_i n_b)}{1 + \sum_{b' \in \mathcal{B}} \exp(v_{ib't} - s_i n_{b'})}, \tag{14}$$

where $\mathbf{X}$ is the matrix of characteristics of all of the products, $\mathbf{p}_t$ is the vector of prices of all of the products and n_b and $n_{b'}$ correspond to the number of supermarkets visited to purchase baskets b and b', respectively.

4.2.3 Restricting to the researcher's choice set

A typical consumer purchases many different products and visits more than one store per month. Estimating a demand model with the full choice entails a dimensionality problem.[9] To keep the problem involving multiproduct and multistore choices tractable, restrictions need to be imposed on the consumers' choice set used for demand estimation.

Assume that the researcher considers only a subset of products and stores, and includes them in a list l. Let $\mathcal{B}_l$ denote the set of baskets composed exclusively of listed products and listed stores. Next, denote $\mathcal{B}_{nl}$ the set of baskets composed of exclusively non-listed products and non-listed stores. Its complement set, $\mathcal{B}_{nl}^c$, contains the set of baskets composed exclusively of listed products and listed stores, $\mathcal{B}_l$, as well as the set of all baskets composed of at least one non-listed product-store with a listed product-store, which is characterised by the cartesian product $\mathcal{B}_l \times \mathcal{B}_{nl}$. The set of baskets $\mathcal{B}$ can characterized by $\mathcal{B}_l \cup \mathcal{B}_{nl} \cup \mathcal{B}_l \times \mathcal{B}_{nl}$.

Consider a basket b^I that contains a combination of listed products and listed stores such

[9] Here we assume that all purchases and visits to stores are observed. Under this setting, an approach to deal with large choice sets is proposed by McFadden (1978)'s and consists on estimating MNL models from subsets of the observed large choice sets. Although, our approach consists on restricting the researcher's choice set, our approach differs substantially from McFadden's approach is so that we exploit the information from the entire choice set. Moreover, our approach allows for unobserved heterogeneity, while McFadden (1978)'s approach does not.

that $b^I \subseteq \mathcal{B}_l$. The probability of purchasing a basket containing b^I is given by the sum over the choice probabilities over all baskets containing b^I, which yields:

$$Q_{ib^It}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = P_{ib^It} + \sum_{h \in b^I \times \mathcal{B}_{nl}} P_{iht} = \frac{\exp(v_{ib^It} - n_{b^I}s_i) + \sum_{h \in b^I \times \mathcal{B}_{nl}} \exp(v_{iht} - n_h s_i)}{1 + \sum_{j \in \mathcal{B}} \exp(v_{ijt} - n_j s_i)}, \quad (15)$$

where n_{b^I} (n_h, n_j or n_g) is the number of stops needed to purchase all of the goods in the basket b^I (h, j, or g). We re-express the denominator in terms of baskets of non-listed product-stores and its complement set:

$$Q_{ib^It}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(v_{ib^It} - n_{b^I}s_i) + \sum_{h \in b^I \times \mathcal{B}_{nl}} \exp(v_{iht} - n_h s_i)}{1 + \sum_{j \in \mathcal{B}_{nl}} \exp(v_{ijt} - n_j s_i) + \sum_{g \in \mathcal{B}^c_{nl}} \exp(v_{igt} - n_g s_i)}. \quad (16)$$

Next, as the first term in the numerator is contained in the second term, we factor the numerator by the first term yielding:

$$Q_{ib^It}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(v_{ib^It} - n_{b^I}s_i) \left[1 + \sum_{k \in \mathcal{B}_{nl}} \exp(v_{ikt} - n_{(k,b^I)} s_i)\right]}{1 + \sum_{j \in \mathcal{B}_{nl}} \exp(v_{ijt} - n_j s_i) + \sum_{g \in \mathcal{B}^c_{nl}} \exp(v_{igt} - n_g s_i)}, \quad (17)$$

where $n_{(k,b^I)}$ denotes the number of additional stops needed to purchase the non-listed goods conditional on purchasing b^I. For any basket $g \subseteq \mathcal{B}_l$, we define Γ_{2igt} as:

$$\Gamma_{2igt} = 1 + \sum_{k \in \mathcal{B}_{nl}} \exp(v_{ikt} - n_{(k,g)} s_i),$$

where $n_{(k,g)}$ denotes the number of additional stops needed to purchase the additional goods conditional on purchasing basket g. Using this expression, we re-express the denominator as follows:

$$Q_{ib^It}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(v_{ib^It} - n_{b^I}s_i)\Gamma_{2ib^It}}{1 + \sum_{j \in \mathcal{B}_{nl}} \exp(v_{ijt} - n_j s_i) + \sum_{g \in \mathcal{B}_l} \exp(v_{igt} - n_g s_i)\Gamma_{2igt}}. \quad (18)$$

Next, by denoting the first two elements of the numerator of equation (18) as:

$$\Gamma_{1it} = 1 + \sum_{j \in \mathcal{B}_{nl}} \exp(v_{ijt} - n_j s_i),$$

we can re-expressed (18) as:

$$Q_{ib^It}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(v_{ib^It} - n_{b^I}s_i)\Gamma_{2ib^It}}{\Gamma_{1it} + \sum_{g \in \mathcal{B}_l} \exp(v_{igt} - n_g s_i)\Gamma_{2igt}}. \quad (19)$$

Taking Γ_{1it} as common factor in the denominator and then re-arranging the terms to be included within the exponential, we obtain that the probability of purchasing a basket containing b^I is given by:

$$Q_{ib^It}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(v_{ib^It} - n_{b^I} s_i - ln(\Gamma_{1it}) + ln(\Gamma_{2ib^It}))}{1 + \sum_{g \in \mathcal{B}_l} \exp(v_{igt} - n_g s_i - ln(\Gamma_{1it}) + ln(\Gamma_{2igt}))}. \tag{20}$$

A simple example. To better understand the relationship between Γ_{1it} and Γ_{2ib^It}, consider the following example. Let a set of choices consisting of three product-store goods A, B, and C and the "no purchase option". Assume that all baskets made of combinations of products A and B make the set $\mathcal{B}_l$, while the set $\mathcal{B}_{nl}$ contains the purchase of only product C. The term Γ_{1it}, equals:

$$\Gamma_{1it} = 1 + \exp(\overline{v}_{ict} - n_l s_i), \tag{21}$$

while the second term, Γ_{2ib^It}, for any $b^I \subseteq \mathcal{B}_l$, equals:

$$\Gamma_{2ib^It} = 1 + \exp(\overline{v}_{ict} - n_{(k,b^I)} s_i). \tag{22}$$

The main difference between Γ_{1it} and Γ_{2it} comes from the term n_l and $n_{(k,b^I)}$. If the good C is located in the same store as A (or B), then $\Gamma_{1it} \neq \Gamma_{2ib^It}$, otherwise if the good is located on a different store than A or B then $\Gamma_{1it} = \Gamma_{2ib^It}$. This captures the idea that, unless the utility of the bundle is super-modular (or sub-modular) with respect to utilities of the single product-stores the two terms, Γ_{1it} and Γ_{2ib^It}, will cancel themselves out.

Extension to multiple units. Consider now a choice set consisting of three product-store goods A, B, and C and the "no purchase option". The maximum number of units that each individual may purchase of each good is given by M. Assuming that all combinations of products A and B make the set $\mathcal{B}_l$. The probability of observing the purchase of product-store good A is given by:

$$Q_{iat}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(\tilde{v}_{iat} - s_i - ln(\Gamma_{1it}) + ln(\Gamma_{2iat}))}{1 + \sum_{g \in \mathcal{B}_l} \exp(\tilde{v}_{igt} - n_g s_i - ln(\Gamma_{1it}) + ln(\Gamma_{2igt}))}, \tag{23}$$

where,

$$\Delta_{ikt} = 1 + \sum_{m=1}^{M} \exp(m \times \overline{v}_{ikt}), \;\; \tilde{v}_{iat} = \overline{v}_{iat} + ln(\Delta_{iat}) - \tau D_{ia}, \;\text{ and}$$
$$\tilde{v}_{igt} = \sum_{k \in g} \overline{v}_{ikt} + ln(\Delta_{ikt}) - \tau D_{ik}.$$

This implies that focusing on a single purchase of product-store goods given that individuals by more than one unit may bias the results by $ln(\Delta_{ikt})$. Provided that the maximum number of units that the consumer can purchase per product-store good is known, then it is possible to include $ln(\Delta_{ikt})$ in the estimation.

4.3 Estimation approach

Individual full purchase history (FPH). We assume that individuals' choice sets are potentially heterogeneous across i's but stable over the choice situations. For each individual in our data, we assume that its choice set is composed of the collection of all the alternatives that individual i is observed to choose in any of the T situations. The choice set is composed of the individual's full purchase history (Crawford, Griffith and Iaria, 2020). That is, we identify consumers that never purchase particular products over the long time dimension of our data as having zero probability of purchasing those products (Dubois, O'Connell and Griffith, 2020). Let each individual FPH choice set be denoted by $\mathcal{B}_i$. To remain consistent with our previous notation, we denote $\mathcal{B}_{li}$ ($\mathcal{B}_{nli}$) the set of all possible products-store combinations with the subset considered by the researcher (the complementary set). The choice probability for basket b^I equals the sum over the choice probabilities over all baskets containing b^I, which yields:

$$Q_{ib^I t}(\boldsymbol{X}, \mathbf{p}; \boldsymbol{\theta}) = \frac{\exp(v_{ib^I t} - n_{b^I} s_i - ln(\Gamma_{1it}) + ln(\Gamma_{2ib^I t}))}{1 + \sum_{j \in \mathcal{B}_{li}} \exp(v_{ijt} - n_j s_i - ln(\Gamma_{1it}) + ln(\Gamma_{2ijt}))}, \tag{24}$$

where Γ_{1it} and $\Gamma_{2ib^I t}$ are defined over $\mathcal{B}_{nli}$.

Estimation. To estimate the parameters of our model, we use the data set described in Section 3. Given that we allow for random coefficients of price and shopping costs, our choice probabilities do not have a closed-form solution. Thus, we use simulated methods to estimate them. In our data, we observe consumers choosing a basket of products (which may be the outside option) at each period t during a number of T periods. Let H index the set of all possible sequences of choices our data takes; that is, all basket sequences at all choice occasions during our period of observation. The probability of observing consumer i making a sequence of choices $h \in H$ is given by:

$$\mathcal{P}_{ih}(\mathbf{X}, \mathbf{p}; \boldsymbol{\theta}) = \prod_{t=1}^{T} Q_{ib^I t}(\mathbf{X}, \mathbf{p}_t; \boldsymbol{\theta}). \tag{25}$$

Let $\mathbf{h}$ be the vector of observed choices of each consumer, a natural way to estimate $\boldsymbol{\theta}$ would be to maximize the log-likelihood function:

$$\mathcal{L}(\mathbf{X}, \mathbf{h}; \boldsymbol{\theta}) = \sum_{i} \ln \mathcal{P}_{ih}(\mathbf{X}, \mathbf{p}; \boldsymbol{\theta}). \tag{26}$$

Adjusting by the researcher's choice set. Let P_{iot} denote individual's i probability of no purchase at week t, and $P_{ib^I t}$ denote the probability that only the goods within basket b^I are purchased by individual i. We can re-express Γ_{1it} as

$$\Gamma_{1it} = 1 + \sum_{j \in \mathcal{B}_{nli}} \exp(v_{ijt} - n_j s_i) = \frac{Q_{iot}}{P_{iot}},$$

and, $\Gamma_{2ib^I t}$ as

$$\Gamma_{2ib^I t} = 1 + \sum_{k \in \mathcal{B}_{nli}} \exp(v_{ikt} - n_{(k,b^I)} s_i) = \frac{Q_{ib^I t}}{P_{ib^I t}}.$$

Given that it is not possible to observe Γ_{1it} for an individual i on each period t, we approximate Γ_{1it} using the ratio between each individuals observed probabilities Q_{iot} and P_{iot} averaged over periods for each consumer yielding $\overline{\Gamma}_{1i} = \overline{Q}_{io}/\overline{P}_{io}$. We approximate the second term using the ratio between each individuals observed probabilities of purchasing any of the baskets that contain listed product and stores averaged over periods, $\overline{Q}_{ib^I}$, with the observed probability of purchasing any of the baskets composed exclusively of listed products averaged over periods $\overline{P}_{ib^I}$ yielding $\overline{\Gamma}_{2i} = \overline{Q}_{ib^I}/\overline{P}_{ib^I}$.

4.4 Monte Carlo simulations

The full choice set. Our demand model allows shoppers to buy several different products in the same week, and assumes that shoppers are making a series of multiple-discrete decisions regarding which products to buy as part of a desired basket of products from a set of mutually exclusive and exhaustive alternatives. We consider a simple model with a full choice set containing three product categories, three stores and a no purchase option. Any combination between the different products and stores is allowed. We assume that household purchase one unit per period of each product, which limits the choice set to 512 ($= 2^9$) alternatives. We simulate the choices for 80 consumers.

The restricted choice sets. For illustration purposes, we consider two scenarios. The first scenario sets an empirical choice set of 16 baskets composed from all combinations of four listed product-store goods ($= 2^4$) and a 'new' outside option. The four goods are described in Table 7. The 'new' outside option is defined as all baskets not including any of the listed products along with the no purchase occasion. To separately identify the shopping costs, we include product-stores from two different stores. We proceed with our estimation and report results in Panel A of Table 7.

As the researchers choice set can be set arbitrarily small, in the second scenario we restrict the choice set to one product-store choice, a singleton basket, and a 'new' outside option (i.e., not purchasing the sole listed good). To identify all product-store preferences in a single estimation, we constructed 9 of such restricted choice set (i.e, one for each of the products-store goods) and stack them together. Finally, to separately identify the shopping costs from the product-store preferences, we stack the choice set of first scenario to the second. For each individual, each choice occasion consists of $9 \times 2 + 16$ observations. We jointly estimate the preferences and report results in Panel B of Table 7.

Shopping period and adjustments. We assume that each shopping period a household makes the decision to purchase or not grocery products. We set the number of periods to 208

(= 4×52), which is consistent with the number of periods in our Nielsen consumer panel data set. To compute the empirical adjustments to correct for the restricted choice set, we approximate Γ_{1it} using the ratio between each individuals observed probabilities Q_{iot} and P_{iot} averaged over the 208 periods for each consumer yielding $\overline{\Gamma}_{1i} = (1 - \overline{Q}_{ib^I})/\overline{P}_{io}$. We approximate the second term using the ratio between each individuals observed probabilities of purchasing any of the baskets that contain included product and stores averaged over periods, $\overline{Q}_{ib^I}$, with the observed probability of purchasing any of the included baskets (and only those baskets) averaged over all 208 periods $\overline{P}_{ib^I}$ yielding $\overline{\Gamma}_{2i} = \overline{Q}_{ib^I}/\overline{P}_{ib^I}$.

Parameter values. The first column of Table 7 reports on the parameter values underlying the true preferences. The product-store preferences are all positive and range between 0.5 and 3. Shopping costs and prices dis-utility are assumed to have a mean value of 3. For each, we assume that the individual heterogeneity follows a normal distribution with a mean value of zero, and a standard deviation of 2. We assume both distributions are independent and identically distributed both within individuals and across individuals. Once an individual's realization is drawn it does not change between periods.

Results. Table 7 reports on the average coefficients and the 5^{th} and 95^{th} percentile (in brackets) from 50 Monte Carlo Simulations for each of the two scenarios considered. The first column reports on the true parameter values used to simulate individual choices for each of the 80 consumers. Panel A reports on results from restricting the choice sets to product 1 and 3 from store 1, and product 1 and 3 from store 2, while Panel B reports on the results for all products. Estimates for mean product preferences in panel A are for both standard and adjusted models are similar. The confidence intervals for the standard model, however, does not contain the true model parameters for 2 out of 4 parameters, while adjusted model's confidence intervals always contains them. A similar finding is reported in Panel B where 7 out of 9 confidence intervals from the standard model do not contain the true parameters, while with the adjusted model all of them are held within its confidence intervals.

For both panel A and panel B, the mean shopping costs and its standard deviation are always underestimated in the standard model and the true parameter values are not contained in the confidence intervals. The adjusted model's shopping costs and standard deviation, however, are correctly estimated and the true values are always contained within the confidence intervals. The price dis-utility coefficient and its standard deviation are correctly estimated by both standard and adjusted models.

5 Conclusions

[To be completed]

Table 7: Estimates for the utility parameters and shopping costs[a]

	True model	Panel A Standard	Panel A Corrected	Panel B Standard	Panel B Corrected
Mean preferences					
product 1 - store 1	**3**	3.09	3.05	2.68	3.06
		[3.01 , 3.24]	[2.97 , 3.20]	[2.58 , 2.80]	[2.96, 3.19]
product 2 - store 1	**2**			2.30	1.89
				[2.19 , 2.43]	[1.68 , 2.06]
product 3 - store 1	**1**	1.07	1.05	1.10	1.08
		[1.01 , 1.12]	[0.98 , 1.10]	[1.05 , 1.20]	[0.99 , 1.17]
product 1 - store 2	**1**	0.87	0.91	0.63	1.03
		[0.79 , 0.96]	[0.83 , 1.01]	[0.53 , 0.75]	[0.93 , 1.12]
product 2 - store 2	**0.5**			0.32	0.63
				[0.18 , 0.42]	[0.48 , 0.77]
product 3 - store 2	**2**	1.72	1.79	1.30	1.86
		[1.64 , 1.81]	[1.71 , 1.88]	[1.22 , 1.40]	[1.77 , 2.01]
product 1 - store 3	**1**			0.40	0.86
				[0.32 , 0.52]	[0.75 , 1.01]
product 2 - store 3	**1**			0.48	0.97
				[0.35 , 0.62]	[0.68 , 1.15]
product 3 - store 3	**0.5**			0.03	0.46
				[-0.09 , 0.15]	[0.37 , 0.58]
Shopping costs					
mean	**3**	**2.32**	**2.81**	**1.70**	**2.95**
		[2.13 , 2.64]	[2.61 , 3.17]	[1.45 , 1.94]	[2.70 , 3.25]
standard deviation	**1**	**0.88**	**1.05**	**0.83**	**1.06**
		[0.75 , 1.01]	[0.86 , 1.23]	[0.63 , 1.02]	[0.88 , 1.22]
Price disutility					
mean	**4**	4.16	4.15	4.09	4.11
		[3.67 , 4.67]	[3.67 , 4.66]	[3.60 , 4.46]	[3.62 , 4.48]
standard deviation	**2**	1.96	1.95	1.92	1.92
		[1.51 , 2.31]	[1.50 , 2.30]	[1.63 , 2.24]	[1.63 , 2.24]

Notes: Results are based on 50 simulations from 80 consumers, each with 208 choice occasions. Full choice set includes 512 baskets, while the IFPH includes on average 20 baskets. Shopping costs and price dis-utility unobserved heterogeneity follow a normal distribution. Brackets represent the 95% confidence intervals.

References

Aguiar, Mark, and Erik Hurst. 2007. "Life-cycle prices and production." *The American Economic Review*, 97(5): 1553–1559.

Anderson, Rod, and Ronald Johnson. 1999. "Antitrust and Sales-Below-Cost Laws: The Case of Retail Gasoline." *Review of Industrial Organization*, 14.

Brief, Richard. 1967. "Measuring Shopping Costs." *Journal of Industrial Economics*, 15(3): 237–241.

Chen, Zhijun, and Patrick Rey. 2012. "Loss leading as an exploitative practice." *The American Economic Review*, 102(7): 3462–3482.

Chen, Zhijun, and Patrick Rey. 2019. "Competitive Cross-Subsidization." *RAND Journal of Economics*, 50(3): 645–665.

Chevalier, Judith, Anil Kashyap, and Peter Rossi. 2003. "Why don't prices rise during periods of peak demand? Evidence from scanner data." *American Economic Review*, 93(1): 15–37.

Crawford, Gregory, Rachel Griffith, and Alessandro Iaria. 2020. "A survey of preference estimation with unobserved choice set heterogeneity." *Journal of Econometrics.*

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform pricing in U.S. retail chains." *Quarterly Journal of Economics*, 134: 2011–2084.

De los Santos, Babur, Ali Hortaçsu, and Matthijs Wildenbeest. 2012. "Testing models of consumer search using data on web browsing and purchasing behavior." *The American Economic Review*, 102(6): 2955–2980.

Dolfen, Stefano, Liran Einav, Peter Klenow, Benjamin Klopack, Jonathan Levin, Larry Levin, and Wayne Best. 2022. "Assessing The Gains from E-Commerce." *American Economic Journal: Macroeconomics*, Forthcoming.

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

Dubois, Pierre, and Helena Perrone. 2015. "Price Dispersion and informational frictions: Evidence from supermarket purchases."

Dubois, Pierre, Martin O'Connell, and Rachel Griffith. 2020. "How well targeted are soda taxes?" *American Economic Review*, 110.

Ellison, Glenn. 2005. "A model of add-on pricing." *Quarterly Journal of Economics*, 120(2): 585–637.

Ellison, Glenn, and Sara Fisher Ellison. 2009. "Search, obfuscation, and price elasticities on the internet." *Econometrica*, 77(2): 427–452.

Ershov, Daniel, Jean-William Laliberté, Mathieu Marcoux, and Scott Orr. 2021. "Estimating complementarity with large choice sets: an application to mergers." *Unpublished manuscript.*

Florez-Acosta, Jorge, and Daniel Herrera-Araujo. 2020. "Multiproduct retailing and consumer shopping behavior: The role of shopping costs." *International Journal of Industrial Organization*, 68.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded attributes, consumer myopia and information supression in competitive markets." *Quarterly Journal of Economics*, 121(2): 505–540.

Gagnon, Etienne, and David López-Salido. 2020. "Samll price responses to large demand shocks." *Journal of the European Economic Association*, 18(2): 792–828.

García-Lebergman, Ezequiel. 2021. "Multi-establishment Firms, Pricing and the Propagation of Local Shocks: Evidence from US Retail." *Unpublished manuscript.*

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

Gerstner, Eitan, and James Hess. 1990. "Can bait and switch benefit consumers?" *Marketing Science*, 9.

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

Hess, James, and Eitan Gerstner. 1987. "Loss leader pricing and rain check policy." *Marketing Science*, 6(4): 358–374.

Hitsch, Günter, Ali Hortaçsu, and Xiliang Lin. 2019. "Prices and Promotions in U.S. Retail Markets: Evidence from Big Data." *NBER Working Paper No. 26306*, 91p.

Holton, Richard. 1957. "Price discrimination at retail: the supermarket case." *Journal of Industrial Economics*, 6(1): 13–32.

Hong, Han, and Matthew Shum. 2006. "Using price distributions to estimate search costs." *RAND Journal of Economics*, 37(2): 257–275.

Honka, Elizabeth. 2014. "Quantifying search and switching costs in the U.S. auto insurance industry." *RAND Journal of Economics*, 45(4): 847–884.

Hortaçsu, Ali, and Chad Syverson. 2004. "Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 Index Funds." *The Quarterly Journal of Economics*, 119(2): 403–456.

Iaria, Alessandro, and Ao Wang. 2020. "Identification and estimation of demand for bundles." *Unpublished manuscript.*

Johansen, Bjorn. 2000. "Demand, Information, and competition: Why do food prices fall at seasonal demand peaks?" *Journal of Industrial Economics*, 48(1): 27–45.

Johnson, Justin. 2017. "Unplanned purchases and retail competition." *American Economic Review*, 107(3): 931–965.

Kim, Jun, Paulo Albuquerque, and Bart Bronnenberg. 2010. "Online demand under limited consumer search." *Marketing Science*, 29(6): 1001–1023.

Klemperer, Paul. 1992. "Equilibrium product lines: competing head-to-head may be less competitive." *The American Economic Review*, 82(4): 740–755.

Koulayiev, Sergei. 2014. "Search for differentiated products: identification and estimation." *RAND Journal of Economics*, 45(3): 553–575.

Lal, Rajiv, and Carmen Matutes. 1994. "Retail pricing and advertising strategies." *Journal of Business*, 67(3): 345–370.

Lazear, Edward. 1995. "Bait and switch." *Journal of Political Economy*, 103(4): 813–830.

Leung, Justin, and Zhonglin Li. 2021. "Big-box store expansion and consumer welfare." *Working Paper*, 67 p.

Moraga-Gonzalez, Jose Luis, Zsolt Sandor, and Matthijs Wildenbeest. 2013. "Semi-Nonparametric estimation of consumer search costs." *Journal of Applied Econometrics*, 28(7): 1205–1223.

Peltier, James, Mark Skidmore, and George Milne. 2013. "Assessing the Impact of Gasoline Sales-Below-Cost Laws on Retail Price and Market Structure: Implications for Consumer Welfare." *Journal of Public Policy and Marketing*, 32(2): 239–254.

Shy, Oz. 2002. "A quick-and-easy method for estimating switching costs." *International Journal of Industrial Organization*, , (20): 71–87.

Simester, Duncan. 1995. "Signalling price image using advertised prices." *Marketing Science*, 14.

Skidmore, Mark, James Peltier, and James Alm. 2005. "Do state motor fuel sales-below-cost laws lower prices?" *Journal of Urban Economics*, 57.

Thomassen, Oyind, Howard Smith, Stephan Seiler, and Pasquale Schiraldi. 2017. "Multi-Category Competition and Market Power: A model of Supermarket Pricing." *American Economic Review*, 107(8): 2308–2351.

Verboven, Frank. 1999. "Product line rivalry and market segmentation–with an application to automobile optional engine pricing." *The Journal of Industrial Economics*, 47(4): 339–425.

Viard, Brian. 2007. "Do switching costs make markets more or less competitive? The case of 800-number portability." *RAND Journal of Economics*, 38(1): 146–163.

Walters, Rockney. 1988. "Retail promotions and retail store performance: a test of some key hypotheses." *Journal of Retailing*, 64.

Walters, Rockney, and Scott MacKenzie. 1988. "A Structural Equations Analysis of the Impact of Price Promotions on Store Performance." *Journal of Marketing Research*, 25(1): 51–63.

Wang, Ao. 2021. "A BLP Demand Model of Product-Level Market Shares with Complementarity." *Unpublished manuscript.*

Warner, Elizabeth, and Robert Barsky. 1995. "The timing and magnitude of retail stores markdowns: evidence weekends and holidays." *Quarterly Journal of Economics*, 110(2): 321–352.

Wildenbeest, Matthijs. 2011. "An empirical model of search with vertically differentiated products." *RAND Journal of Economics*, 42(4): 729–757.