Personalized Pricing and the Value of Past Purchase Histories: An Empirical Perspective*

Isis Durrmeyer† Jean-François Fournel‡ Mario Samano§

October 30, 2024

Abstract

Our analysis uses data on prices, aggregate quantities, and individual purchase histories from a large supermarket chain in the US and an empirical model to represent grocery shopping by consumers and the supermarket pricing strategies. We estimate demand for 23 product categories and supermarket marginal costs consistent with the observed uniform price setting. With the estimated distribution of preferences in hand, we simulate the information acquisition by the supermarket from purchase histories, assuming the supermarket uses Bayes's rule to update its priors about consumers' preferences. We then evaluate how profitable it is to set personalized prices using the information contained in purchase historie and the consequences for consumer surplus. Our results show that the supermarket can reap between 60% to 80% of the potential gains from perfect price discrimination, depending on the product category. Aggregate consumer surplus decreases and at the individual level, we find very asymmetric price effects. Prices increase a lot for a small fraction of individuals while we observe small price decrease for a large share of the population.

Keywords: personalized pricing, differentiated goods, demand calibration, price competition.
JEL Classifications: L11, L13, L81

*We acknowledge financial support from the European Research Council under grant ERC-2019-STG-852815 "PRIDISP". We thank the participants at a Toulouse of School of Economics presentation of a preliminary draft for their comments.

†Toulouse School of Economics, Université Toulouse Capitole. E-mail: isis.durrmeyer@tse-fr.eu

‡Toulouse School of Economics, Université Toulouse Capitole. E-mail: jean-francois.fournel@tse-fr.eu

§HEC Montreal. E-mail: mario.samano@hec.ca

1 Introduction

Personalized pricing is a form of price setting that uses explicit market segmentation to approximate perfect price discrimination. Recognizing that such market segmentation is nearly impossible, economists have focused on using consumers' data that may help reveal the consumers' willingness to pay. One particular example of such data is past purchase information. Firms set prices based on the sequence of past purchase histories, a pricing strategy known as behavior-based price discrimination (Fudenberg and Villas-Boas, 2006).

In many product markets, firms record consumers' purchases. Firms offer loyalty programs, which typically provide consumers with benefits. In exchange, firms can track individual purchases over time. Creating an account to purchase in online markets is often necessary; a side effect is that purchases are recorded. Additionally, firms nowadays have better technologies to set individualized prices. Firms can offer specific promotions and discounts to some consumers and condition their values on past purchase histories. For instance, supermarkets use apps to send promotional discounts to consumers. A concrete example is from more than ten years ago: some supermarkets already implemented different forms of personalized pricing on goods such as bottled water.[1] More recently, with the advent of online platforms, the cost of acquiring information has decreased, and new methods have been developed to offer personalized prices to potential buyers.[2] This type of business case has raised interest from regulators in the European Union and the Federal Trade Commission.[3]

In this paper, we investigate to what extent the information collected by sellers on consumers' purchases over time can be used to improve profits using behavior-based price discrimination and what the consequences are for consumers. The fundamental economic question behind the profitability of behavior-based price discrimination is the informational content of past purchase histories. How much can firms learn about consumers' preferences from observing their previous purchases? And what are the consequences for prices and consumers?

We develop a structural model where a multi-product seller knows the overall distribution of consumer preferences and learns from consumers' purchase histories by applying Bayes's rule. Then, the posterior belief is used to set personalized prices for each consumer.[4] We then

[1] Safeway (2012) `https://t.ly/k9dLm`.

[2] Ziprecruiter (2018) `https://t.ly/N2dmA`.

[3] European Union: `https://t.ly/z1A_o`, the OECD and the FTC `https://t.ly/q7Ueb.`

[4] We use interchangeably the terms *personalized pricing*, *targeted pricing* and *behavior-based price discrimination*.

apply this model to grocery purchases to quantify the welfare consequences of behavior-based price discrimination. The quantification of the value of past purchase data in terms of profits and consumer surplus when these data are exploited to personalize prices offers an objective framework to reflect on these business strategies. This speaks directly to the open debate regarding consumer protection data and the monetization of purchase occasions records.

We use a dataset from a U.S. supermarket that covers 23 product categories, spans over 12 months, and includes the purchase histories of 17,756 consumers. We first estimate the distribution of consumer preferences in each product category. We combine aggregate sales and repeated individual purchases to estimate the distribution of price sensitivities using a likelihood function with constraints. The likelihood function describes the consumer's joint purchase histories. This estimation method is in the spirit of traditional likelihood-based demand estimation routines such as that in Goolsbee and Petrin (2004) and, more recently, in Grieco et al. (2022). We estimate several different demand systems, each representing a product category. There are multiple examples in the literature on demand estimation using supermarket data (Rossi et al., 1996, Nevo, 2001, Thomassen et al., 2017, Eizenberg et al., 2021, Smith et al., 2022, to mention a few). However, a salient point of our analysis is the use of several distinct categories of products. A recent study that also estimates demand for a large set of product categories is Döpper et al. (2023).

Once we have calibrated the demand functions, we assume that the supermarket takes the wholesale prices as given and sets prices uniformly for all consumers. Therefore, we model the supermarket as a multi-product monopolist for each product category, and we can recover marginal costs from the first-order conditions associated with profit maximization. We estimate marginal costs that are very consistent with the observed wholesale prices.[5] We take this as a robust sign that our model accurately represents the market structure.

We then simulate new market equilibria under a regime with behavior-based price discrimination using purchase history. To do this, we assume that the firm observes consumers' purchase decisions during several periods and uses this information to form a belief about consumer types using Bayes' rule. Specifically, the supermarket determines the probability distribution of consumer types conditional on the realized past purchases. There have been other studies that compute personalized prices using supermarket data (Rossi et al., 1996 and Shiller, 2020, for instance). In comparison, we explicitly model the supply side and capture

[5]Even though we have data on wholesale prices, these are not used in the calibration. Instead, we use them only to assess the model's goodness of fit.

business-stealing effects from under or over-pricing relative to the uniform price. In addition, we consider a case in which the supermarket has perfect information on the consumer types and performs first-degree price discrimination.

Once we implement the personalized prices based on purchase histories, we observe that consistent with economic intuition, profits increase, and consumer surplus decreases. The supermarket reaps between 60 and 80% of the profits it would gain under perfect price discrimination. The gains monotonically increase with the purchase history's length and are concave with history length, indicating decreasing marginal returns to information from purchase histories.

On the consumer side, we find that some consumer types are presented with supra-uniform prices while others with infra-uniform prices. However, the shares of consumers in each of those groups are not symmetric. The share of the most price-elastic consumers that receive an infra-uniform price is about 79 to 96%, depending on the product category. Despite that, overall consumer surplus decreases relative to the uniform price benchmark. This stems from the fact that the maximum discount offered to price-sensitive consumers is 25.4%, while the maximum surcharge imposed on price-insensitive consumers is 195%.

Related literature. Our demand and supply models are based on the classical setup by Berry (1994) and Berry et al. (1995). Most of the literature listed below also uses the same discrete choice setup.

Personalized pricing under monopoly has been implemented using Bayesian updating since Rossi et al. (1996) (five categories of canned tuna, no supply side); by Shin et al. (2012); and more recently by Dubé and Misra (2023) who rely also on experimental data. Other approaches in the monopolistic case include Waldfogel (2015).

The case of personalized pricing under competition has been studied by Dubé et al. (2017) from an empirical and computational perspective and by Chen et al. (2020), Rhodes and Zhou (2022), Ali et al. (2022), and Choe et al. (2022) from a theory perspective in horizontal market structures. With vertical market structures, Jullien et al. (2023) show that it is possible to decrease the negative effects of competition on firms when there is personalized pricing.

Empirical evidence of the extent that supermarkets implement uniform pricing include DellaVigna and Gentzkow (2019) (evidence of uniform pricing in US supermarkets), Hitsch et al. (2019) (weaker evidence of uniform pricing), Chandra and Lederman (2018), Puller and Taylor (2012), and Eizenberg et al. (2021) (different prices in different neighborhoods in a city).

Albeit different from our approach to calibrate demand systems, demand calibration has been used in Miller et al. (2016) who studied the implications of different functional forms (logit, linear, log-linear, and AIDS) to simulate mergers. Miller et al. (2013) show that data on pass-throughs can be used to calibrate demand systems as well.

Some examples of demand estimation using supermarket product data include Döpper et al. (2023) (100 distinct product categories assuming profit maximization and data on prices and quantities (no observed costs)), Compiani (2022) (non-parametric, strawberries), Thomassen et al. (2017) (multi-store multi-category model of consumer demand), and Smith et al. (2022) (mayonnaise, 3 brands).

The rest of the paper is organized as follows. Section 2 describes our theoretical framework under alternative pricing strategies. Section 3 addresses the calibration of the demand systems. Section 4 presents the data and the calibration results. Sections 5 and 6 present the core results from our simulation exercise. Section 7 provides some concluding remarks.

2 The model

Demand. We consider demand arises from a heterogeneous set of consumers. First, we consider individuals have random purchase occasions. The probability of receiving a purchase occasion is uniform across individuals and exogenous. We denote by ρ this probability of having a purchase occasion, i.e. to enter the supermarket. Individuals are heterogeneous in price sensitivity but have the same mean utility of buying each product. The price sensitivity parameter is α, and we assume it follows a distribution represented by the cumulative density function $F_\alpha(\alpha)$ with support $[\underline{\alpha}, \overline{\alpha}]$. We refer to this as the consumer's type.[6] The utility of consumer i buying product j in a purchase occasion is

$$u_{ij} = \delta_j + \alpha_i p_j + \epsilon_{ij},$$

where to keep the number of indices reasonable, we do not write the purchase occasion index. δ_j is product j's specific quality index. There is an outside option with mean utility normalized to 0 such that

$$u_{i0} = \epsilon_{i0}.$$

[6] Alternatively, one could define the consumer type as an additive term or intercept of the utility function. We prefer to define the consumer types by the price coefficient since this explicitly uses our observed data (prices) instead of unobserved product characteristics.

We make the standard assumption that the ϵ_{ij} are identically and independently distributed across individuals and products and follow an extreme value distribution. We also assume that the product taste shocks are independent of the consumer type. The probability that the consumer with the type α purchases product j is

$$s_j(\alpha) = \frac{\exp\left(\delta_j + \alpha p_j\right)}{1 + \sum_{j'} \exp\left(\delta_{j'} + \alpha p_{j'}\right)}.$$

Therefore, the overall market share of product j conditional on a purchase occasion is

$$s_j = \int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha) dF_\alpha(\alpha).$$

Ultimately, the probability of observing a sale of good j is

$$\tilde{s}_j = \rho \int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha) dF_\alpha(\alpha). \tag{1}$$

Consumer surplus. The consumer surplus is the expected utility of the best choice and is given by:

$$CS(\alpha) = -\frac{1}{\alpha} \log\left(1 + \sum_{j=1}^{J} \exp(\delta_j + \alpha p_j)\right)$$

where α is, as previously noted, the price sensitivity parameter. The aggregate consumer surplus is therefore:

$$CS = -\int_{\underline{\alpha}}^{\overline{\alpha}} \frac{1}{\alpha} \log\left(1 + \sum_{j=1}^{J} \exp(\delta_j + \alpha p_j)\right) dF_\alpha(\alpha).$$

Uniform pricing. Throughout, we assume that supermarkets set prices, not manufacturers. Therefore, we consider pricing for a multi-product monopolist for each product category separately. To keep the model tractable, we ignore cross-category interactions and complementarities between products in the supermarket's pricing decision, i.e., prices are set in isolation in each category. There are J_k products in category k. Under uniform pricing, the supermarket sets only one price for all consumers because it cannot price discriminate through personalized pricing (e.g., it does not have the technology) or because it has no information about consumers. Let π_k^{UN} denote the profits from product category k under uniform pricing:

$$\pi_k^{UN} = \left[\sum_{j \in J_k} (p_j - c_j) \int_{\underline{\alpha}}^{\overline{\alpha}} \rho s_j(\alpha) dF_\alpha(\alpha)\right] M,$$

where c_j is the marginal cost of product j and includes the wholesale price and retail costs. M denotes the potential market size. The supermarket sets prices such that the following first-order conditions are satisfied:

$$\mathbf{s} + \Lambda(\mathbf{p} - \mathbf{c}) = 0. \tag{2}$$

where $\mathbf{s}$, $\mathbf{p}$, and $\mathbf{c}$ are the vectors of market shares, prices, and marginal costs, respectively, within category k. Each element of $\mathbf{s}$ is equal to $\int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha) dF_\alpha(\alpha)$. The matrix Λ contains all the derivatives of the market shares such that the element (j, j') of Λ is equal to

$$\Lambda_{jj'} = \begin{cases} \int_{\underline{\alpha}}^{\overline{\alpha}} \alpha s_{j'}(\alpha)(1 - s_{j'}(\alpha)) dF_\alpha(\alpha) & \text{if } j = j', \\ -\int_{\underline{\alpha}}^{\overline{\alpha}} \alpha s_j(\alpha) s_{j'}(\alpha) dF_\alpha(\alpha) & \text{if } j \neq j'. \end{cases}$$

Note that Λ is a function of $\mathbf{p}$ through the market share expressions. The solution to this system of non-linear equations is the vector of uniform prices $\mathbf{p}^{UN}$ for category k.

Firm's information acquisition process. The firm can observe the realizations of the purchase occasions and when a the purchase occasion occurs, the firm observes the choices made by each individual and the price environment. We call the repetition of individual choices for T periods consumers' histories. The firm forms an expectation about consumer's type from the purchase history using Bayes's rule. Denote by $\mathbf{h}_\ell = \{(Y_{\ell 01}, ..., Y_{\ell J1}), ..., (Y_{\ell 0T}, ..., Y_{\ell JT})\}$ the purchase history ℓ of length T, where $Y_{\ell jt}$ is equal to 1 if the consumer buys product j at time period t and 0 otherwise. Note that it is possible that $\sum_{j=0}^{J} Y_{\ell jt}$ equals to 0, when a consumer does not have a purchase occasion. Indeed, there is a key difference between having a purchase occasion and choose not to buy the good and not having a purchase occasion. In particular, the firm is able to infer something about individual preferences in the first case but not in the second case. Note that we use the index ℓ to differentiate it from the index for consumers i because multiple consumers may have the same purchase history. Indeed, if two different consumers have the same sequence of purchases, then we label their histories with the same index. We have a set of L possible purchase histories, $\ell = 1, \ldots, L$. Finally, we denote by $M_{\mathbf{h}_\ell} \geq 0$ the number of potential consumers with purchase history $\mathbf{h}_\ell$.[7] We have $M_{\mathbf{h}_1} + ... + M_{\mathbf{h}_L} = M$. Following Bayes' rule, we can write

$$f_{\alpha|\mathbf{h}_\ell}(\alpha|\mathbf{h}_\ell) = \frac{f_{\mathbf{h}_\ell|\alpha}(\mathbf{h}_\ell|\alpha) f_\alpha(\alpha)}{f_{\mathbf{h}_\ell}(\mathbf{h}_\ell)}. \tag{3}$$

[7]Note that the set of possible histories grows exponentially in T. Therefore, even for moderately short histories we have $L > M$. In this case, some histories ℓ are irrelevant for the firm and are such that $M_{\mathbf{h}_\ell} = 0$.

The probability of observing history h_ℓ conditional of α is:

$$f_{\mathbf{h}_\ell|\alpha}(\mathbf{h}_\ell|\alpha) = \prod_{t=1}^{T}\left[\prod_{j=0}^{J}\left(\rho s_{jt}(\alpha)\right)^{Y_{\ell jt}}\right]\cdot(1-\rho)^{1-\sum_{j=0}^{J}Y_{\ell jt}}$$

because demand shocks are independent across purchase occasions. The unconditional probability of purchase history h_ℓ can be written

$$f_{\mathbf{h}_\ell}(\mathbf{h}_\ell) = \int_{\underline{\alpha}}^{\overline{\alpha}} f_{\mathbf{h}_\ell|\alpha}(\mathbf{h}_\ell|\alpha)dF_\alpha(\alpha).$$

Finally, Equation (3) for the distribution of consumer type conditional on observing purchase history $\mathbf{h}_l$ is

$$f_{\alpha|\mathbf{h}_\ell}(\alpha|\mathbf{h}_\ell) = \frac{\prod_{t=1}^{T}\left[\prod_{j=0}^{J}\left(\rho s_{jt}(\alpha)\right)^{Y_{\ell jt}}\right]\cdot(1-\rho)^{1-\sum_{j=0}^{J}Y_{\ell jt}}f_\alpha(\alpha)}{\int_{\underline{\alpha}}^{\overline{\alpha}} f(\mathbf{h}_\ell|\alpha)dF_\alpha(\alpha)}.$$

Personalized pricing. Under personalized pricing, the supermarket sets prices $\mathbf{p}(\mathbf{h}_\ell)$ for all products in category k, conditionally on observing purchase histories $\mathbf{h}_\ell$. The profits function can be expressed as

$$\pi_k^{PS} = \sum_{\ell=1}^{L}\sum_{j\in J_k}\left(p_j(\mathbf{h}_\ell)-c_j\right)\int_{\underline{\alpha}}^{\overline{\alpha}}\rho s_j(\alpha)dF_{\alpha|\mathbf{h}_\ell}(\alpha|\mathbf{h}_\ell).M_{\mathbf{h}_\ell},$$

where $M_{\mathbf{h}_\ell}$ denotes the number of potential consumers with purchase history $\mathbf{h}_\ell$. The supermarket sets prices for each purchase history separately. The set of optimal prices, one for each history $\mathbf{h}_\ell$, is such that the following system of J_k non-linear first-order conditions is satisfied:

$$\mathbf{s}(\mathbf{h}_\ell) + \Lambda(\mathbf{h}_\ell)\left(\mathbf{p}(\mathbf{h}_\ell)-\mathbf{c}\right) = 0, \tag{4}$$

giving a total of $J_k \times L$ equations. Each element of $\mathbf{s}(\mathbf{h}_\ell)$ is the market share computed using the conditional density obtained through the Bayes's rule above, $s_j(\mathbf{h}_\ell) = \int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha)dF_{\alpha|\mathbf{h}_\ell}(\alpha|\mathbf{h}_\ell)$. Similarly, each element (i,j) of matrix $\Lambda(\mathbf{h}_\ell)$ can be expressed as

$$\Lambda_{jj'}(\mathbf{h}_\ell) = \begin{cases}\int_{\underline{\alpha}}^{\overline{\alpha}}\alpha s_{j'}(\alpha)(1-s_{j'}(\alpha))dF_{\alpha|\mathbf{h}_\ell}(\alpha|\mathbf{h}_\ell) & \text{if } j=j',\\ -\int_{\underline{\alpha}}^{\overline{\alpha}}\alpha s_j(\alpha)s_{j'}(\alpha)dF_{\alpha|\mathbf{h}_\ell}(\alpha|\mathbf{h}_\ell) & \text{if } j\neq j'.\end{cases}$$

Perfect price discrimination. As a benchmark, we take the extreme case where the firm perfectly knows the type of each consumer and can condition its pricing strategy on the type. The profits function is, in this case:

$$\pi_k^{PF} = \sum_{i=1}^{M} \sum_{j \in J_k} \left(p_j(\alpha_i) - c_j\right) \rho s_j(\alpha_i),$$

where M is the number of potential consumers. The supermarket sets prices $\mathbf{p}(\alpha_i)$ for each consumer type separately. The optimal price for a type α_i is such that the following J_k first-order conditions are satisfied,

$$\mathbf{s}(\alpha_i) + \Lambda(\alpha_i)(\mathbf{p}(\alpha_i) - \mathbf{c}) = 0 \tag{5}$$

for each $i = 1, \ldots, M$ and for each category separately.

3 Calibration of the model parameters

We calibrate the demand and cost parameters for each product category separately. Let $k = 1, \ldots, K$ denote the food category. We observe $t = 1, \ldots, T$ periods. First, we estimate ρ the exogenous probability of receiving a purchase occasion.

$$\hat{\rho} = \frac{1}{MT} \sum_{i=1}^{M} \sum_{t=1}^{T} \max \left\{ \sum_{j=0}^{J_1} Y_{ijt}^1, ..., \sum_{j=0}^{J_K} Y_{ijt}^K \right\}$$

We assume the observed market outcomes are the consequence of the retailer setting uniform prices taking the wholesale prices as given. In addition, we assume that the distributions of the price sensitivity are discrete with D_k known points of support. The preference parameters are the type probabilities $\{\phi_d\}_{d=1,\ldots,D_k}$, the mean utilities of products $\{\delta_{jt}\}_{j=1,\ldots,J_k;\, t=1,\ldots,T}$, and the (known) points of support $\{\alpha_d\}_{d=1,\ldots,D_k}$.

We compute the likelihood of observing the joint purchase histories in the data. We observe a sample of M individuals and their respective purchase histories for each product category $\mathbf{h}_i = \{Y_{ijt}\}_{j=0,\ldots,J_k;\, t=1,\ldots,T}$. The contribution of consumer i to the likelihood function is

$$\sum_{d=1}^{D_k} \phi_d^k \prod_{t=1}^{T} \prod_{j=0}^{J_k} s_{jt}(\alpha_d^k, \boldsymbol{\delta}_t^k)^{Y_{ijt}}.$$

The objective function consists of minus the log-likelihood of the sample,

$$G^k(\boldsymbol{\alpha}^k, \boldsymbol{\phi}^k, \{\boldsymbol{\delta}_t^k\}_{t=1,\ldots,T}) = \sum_{i=1}^{M} \log \left(\sum_{d=1}^{D_k} \phi_d^k \prod_{t=1}^{T} \prod_{j=0}^{J_k} s_{jt}(\alpha_d^k, \boldsymbol{\delta}_t^k)^{Y_{ijt}} \right),$$

where the outer sum aggregates over consumers, possibly with some consumers sharing the same purchase history.

We also have a set of constraints for market shares from equation (1),

$$\mathfrak{s}_{jt}^{k} = \hat{\rho} \sum_{d=1}^{D_k} \frac{\exp\left(\delta_{jt}^{k} + \alpha_{d}^{k} p_{jt}\right)}{1 + \sum_{j'=1}^{J_k} \exp\left(\delta_{j't}^{k} + \alpha_{d}^{k} p_{j't}\right)} \phi_{d}^{k}, \quad \forall j, \forall t, \tag{6}$$

where $\mathfrak{s}_{jt}^{k}$ is the observed market share. For each period and for given values of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, this system of equations defines a unique vector of mean utilities $\boldsymbol{\delta_t}$ as shown by Berry (1994). The calibration is based on maximizing G^k, subject to the non-linear constraints given by equation (6) and such that the probabilities of consumer types add up to 1. This calibration method is similar to the MPEC formulation of the BLP model (see, Dubé et al., 2012). Putting everything together, the problem to solve for category k is

$$\begin{aligned} &\max_{\boldsymbol{\phi}^k, \{\boldsymbol{\delta}_t^k\}_{t=1,\dots,T}} \quad G^k(\boldsymbol{\alpha}^k, \boldsymbol{\phi}^k, \{\boldsymbol{\delta}_t^k\}_{t=1,\dots,T}) \\ &\text{s.t.} \quad \mathfrak{s}_{jt}^{k} = \hat{\rho} \sum_{d=1}^{D_k} \frac{\exp\left(\delta_{jt}^{k} + \alpha_{d}^{k} p_{jt}^{k}\right)}{1 + \sum_{j'=1}^{J_k} \exp\left(\delta_{j't}^{k} + \alpha_{d}^{k} p_{j't}^{k}\right)} \phi_{d}^{k}, \quad \forall j, \forall t \\ &\qquad \sum_{d=1}^{D_k} \phi_d^k = 1, \quad \phi_d \geq 0. \end{aligned} \tag{7}$$

4 Data and calibration results

4.1 Data

We calibrate the model using data on food purchases and wholesale prices provided by DecaData. This is a scanner dataset with additional key information. First, the data are at the most disaggregated level, so each observation is a product and a transaction. Second, we observe consumers' loyalty card numbers if they have one so we can track these individuals over time and compute their purchase histories. Third, we have shipping data that contain the wholesale prices of products. Additional information on the data and the composition of the final dataset are available in Appendix B.

We focus on one representative supermarket in 2018. We choose a store near the median store in terms of total sales. Our chosen store has \$13.1 million in sales in 2018, placing it at the 52$^{\text{nd}}$ percentile of the distribution. At the same time, we restrict our analysis to the 17,756 consumers with a fidelity card who made at least one purchase at our representative store in

2018. While this may lead to selection issues, we cannot track the excluded consumers over time and recover their purchase histories, which is crucial for studying personalized pricing. Sales from this subset of consumers totalize \$10.4 million, roughly 79% of all sales recorded at our store.

Due to data limitations, we do not use demographics to reveal consumer types or to compute personalized prices. However, previous studies have concluded that availability on demographics information does not improve profitability whereas information on the timing of purchase occasions does. For instance, in the seminal work by Rossi et al. (1996) demographic variables do not influence the results. Smith et al. (2022) and Shiller (2020) arrive at the same conclusions.

We consider each category to be a separate market and ignore the potential complementarity or substitutability between goods in different categories (e.g., chips and soda are complements as Ershov et al. (2021) point out). We also ignore competition across stores and shopping costs. Thomassen et al. (2017) provide a model to estimate the joint decision of supermarkets and products. However, we do not observe the supermarkets' locations, so we cannot model the competition at the supermarket level. Finally, we focus on monthly purchases and define each month as a separate shopping occasion.[8]

Table 1 summarizes the final sample characteristics. We report the number of products per category (excluding the outside option), the standard package size used to normalize units, and the average monthly consumption (conditional on purchasing) per consumer. We also show the average unit price and cost for one standard unit of the good. Although we report the number of products in each category, recall that we assume that the supermarket sets the final price for consumers, treating each category as a multi-product pricing problem.

4.2 Purchase histories and market shares

We next describe how we recover consumer purchase histories and compute market shares from our data. One difficulty arises from the fact that consumers can purchase positive quantities of multiple goods, while our model allows consumers to purchase at most one unit of one good in each shopping occasion. In cases where more than one product was purchased, we keep the product with the highest absolute quantity purchased in standard units.We formalize this as follows. Recall that we denote consumer i's purchase history of

[8] Consumers visit the store on average 1.52 times per month, see Table B.1.

Table 1: Summary statistics of the final sample.

Category	# Products (J_k)	Standard pack size	Units of measurement	Avg. monthly consumption	Avg. unit price	Avg. unit cost
Alcohol						
Beer	4	144	fl oz	1.85	13.19	10.27
Wine	7	25.4	fl oz	3.73	7.39	5.47
Dairy						
Butter/Margarine/Spread	5	16	oz	2.1	2.29	1.6
Cheese	4	8	oz	3.9	2.55	1.58
Cream/Creamer	4	32	fl oz	1.69	4.15	2.67
Fresh egg	2	12	unit	2.28	2.82	1.94
Refrigerated Juice/Beverage	4	59	fl oz	3.7	2.41	1.71
Milk	5	64	fl oz	4.13	2.65	1.74
Yogurt	4	5.3	fl oz	9	0.79	0.55
Frozen food						
Ice Cream	4	48	fl oz	2.24	4.55	3.03
Frozen Pizza	4	10	oz	5.44	1.73	1.29
Frozen Potatoes/Onions	4	32	oz	2.01	2.87	2.07
Frozen Vegetables	4	12	oz	5.34	1.56	0.98
Grocery						
Cereal	5	12	oz	2.42	3.62	2.94
Coffee	6	12	oz	1.26	10	7.83
Crackers	5	12.4	oz	2.04	3.03	2.4
Snack grocery	3	8	oz	3.93	2.86	2.17
Soft drink/Mixer	5	144	fl oz	4.34	7.65	4.82
Water	6	128	fl oz	5.23	4.13	2.29
Packaged meat						
Bacon	6	24	oz	1.17	9.54	6.78
Dinner sausage	6	14	oz	4.11	2.81	1.81
Taxable grocery						
Bleach/Stain remover	4	128	fl oz	2.59	3.38	2.54
Laundry detergent	4	32	fl oz	2.33	6.61	5.52

NOTES: *Standard Pack Size* is the most common pack size sold in a given category, *Avg. Monthly Consumption* is the average quantity purchased in standard units, conditional on purchasing any good in the category in a given month, *Avg. Unit Price* is the weighted average price for one standardized unit, and *Avg. Unit Cost* is the weighted average cost.

length T as

$$\mathbf{h}_i = \{(Y_{i01}, ..., Y_{iJ1}), ..., (Y_{i0T}, ..., Y_{iJT})\},$$

where the category index k is omitted to simplify the notation, and Y_{ijt} is equal to 1 if product j is chosen in purchase occasion t and 0 otherwise. Let q_{ijt} be the quantity of product j purchased by consumer i in shopping occasion t, where q_{ijt} is expressed in normalized units as per Table 1. We compute the entries in each purchase history as follows:

$$Y_{ijt} = \begin{cases} 1 & \text{if } q_{ijt} \geq q_{ij't}, \forall j \neq j';\ q_{ijt} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1, ..., J_k$ and

$$Y_{i0t} = \begin{cases} 1 & \text{if } q_{ijt} = 0, \forall j \text{ and } \max \left\{ Y_{ijt}^k \right\}_{k=1}^K = 1 \\ 0 & \text{otherwise.} \end{cases}$$

One important requirement in our analysis is that the market shares used in the calibration stage are consistent with the set of histories we use in our simulations. Accordingly, we compute the observed market shares directly from purchase histories, that is

$$s_{jt} = \frac{1}{M} \sum_{i=1}^{M} Y_{ijt}.$$

We assume that the potential market is the set of consumers that made at least one purchase in the store in 2018 while being part of the loyalty program. Finally, we assume that a consumer choosing good j purchases the average monthly quantity instead of one unit of the good, and prices and wholesale prices are modified to reflect this normalization.

4.3 Calibration results

The calibration is performed in two steps. In the first step, we calibrate the model by finding a solution to problem (7) under the assumption that the price sensitivities are normally distributed and estimate the mean and standard deviation α_0 and σ_0 of that distribution. We then use these estimates to construct the support of α as the 90% confidence interval around α_0. In a few cases, the upper bound of the support is positive or very close to 0. In those cases, we replace the upper bound of the support by $\overline{\alpha} = -0.1$.

In the second step, we discretize the support in 50 equal intervals and estimate the discrete probability distribution ϕ_d over the 50 types α_d as the solution to Problem (7). Types with a probability mass below 0.01 are iteratively removed from the support and the calibration is repeated until we have a stable distribution, $\{\alpha_1, ..., \alpha_{D_k}\}$ with probabilities $\{\phi_1, ..., \phi_{D_k}\}$ greater that 0.01.[9]

Table 2 presents the calibration results. Columns 3 and 4 provide the calibrated parameters from the first step that rely on the assumption of normal distributions. Results from the final calibration are presented in columns 5 through 9. We report the number of points on the support that survive our iterative procedure. The number of consumer types varies from

[9] We experimented with lower thresholds of up to 10^{-4} without finding different results overall. Types with a very low probability mass create numerical issues when computing personalized prices.

Table 2: Calibration results

Category	# Products	Parameters of the Normal		Discrete distribution				
	(J_k)	α_0	σ_0	D_k	E(α_d)	SD(α_d)	E($\bar{\delta}_t$)	SD($\bar{\delta}_t$)
Alcohol								
Beer	4	-0.360	0.100	3	-0.503	0.061	3.482	0.313
Wines	7	-0.293	0.099	2	-0.45	0.034	2.643	0.431
Dairy								
Butter/Margarine/Spreads	5	-1.193	0.398	5	-1.486	0.441	0.741	0.211
Cheese	4	-0.496	0.170	5	-0.567	0.175	-0.227	0.119
Creams/Creamers	4	-0.885	0.362	3	-1.418	0.182	2.052	0.476
Fresh eggs	2	-0.993	0.325	5	-1.227	0.319	1.897	0.227
Refrigerated Juice/Beverage	4	-0.711	0.253	6	-0.894	0.272	0.613	0.143
Milk	5	-0.429	0.219	7	-0.582	0.196	-0.383	0.255
Yogurt	4	-0.801	0.285	3	-1.166	0.223	0.945	0.150
Frozen food								
Ice Cream	4	-0.642	0.212	5	-0.844	0.212	0.501	0.619
Frozen Pizza	4	-0.615	0.206	3	-0.898	0.148	0.984	0.164
Frozen Potatoes/Onions	4	-0.941	0.345	5	-1.356	0.269	0.323	0.170
Frozen Vegetables	4	-0.581	0.205	4	-0.719	0.247	-0.427	0.248
Grocery								
Cereal	5	-1.171	0.252	4	-1.388	0.297	1.613	0.273
Coffee	6	-0.766	0.212	3	-1.047	0.162	3.108	0.338
Crackers	5	-1.269	0.304	4	-1.502	0.346	0.531	0.232
Snacks grocery	3	-0.688	0.13	4	-0.735	0.169	2.047	0.150
Soft drinks/Mixers	5	-0.423	0.107	5	-0.489	0.108	1.109	0.482
Water	6	-0.688	0.230	5	-0.927	0.199	1.835	0.438
Packaged meat								
Bacon	6	-0.642	0.191	5	-0.817	0.192	1.380	0.173
Dinner sausage	6	-0.433	0.180	6	-0.571	0.187	-0.165	0.106
Taxable grocery								
Bleach/Stain removers	4	-1.670	0.524	3	-2.331	0.438	2.679	0.812
Laundry detergent	4	-0.802	0.170	3	-1.051	0.106	2.800	0.853

NOTES: α_0 and σ_0 are the mean and standard deviation of the distribution of α assuming a normal distribution. D_k is the number of types recovered when we assume consumers' price sensitivities are distributed on a discrete grid. We report the mean price sensitivity E(α_d) and the standard deviation SD(α_d). We also provide the average mean product valuation E($\bar{\delta}_t$) and the standard deviation SD($\bar{\delta}_t$).

2 to 7, averaging 4.26. Finally, we report the average price sensitivity and the variance of the price sensitivities. In general, the average price sensitivity from the calibration is higher than our initial estimate α_0. This means the calibration assigns more weight to types with a high price sensitivity than the normal distribution. The variance is also slightly lower in most cases, meaning slightly less dispersion in price sensitivity. For completeness, Figure A.1 and Figure A.2 in Appendix A depict the distribution of price sensitivities for each category.

4.4 Goodness of fit

4.4.1 Marginal costs

We recover marginal costs for each product by using the set of first-order conditions (2) for the uniform price case. These reflect the price the supermarket pays to the suppliers.

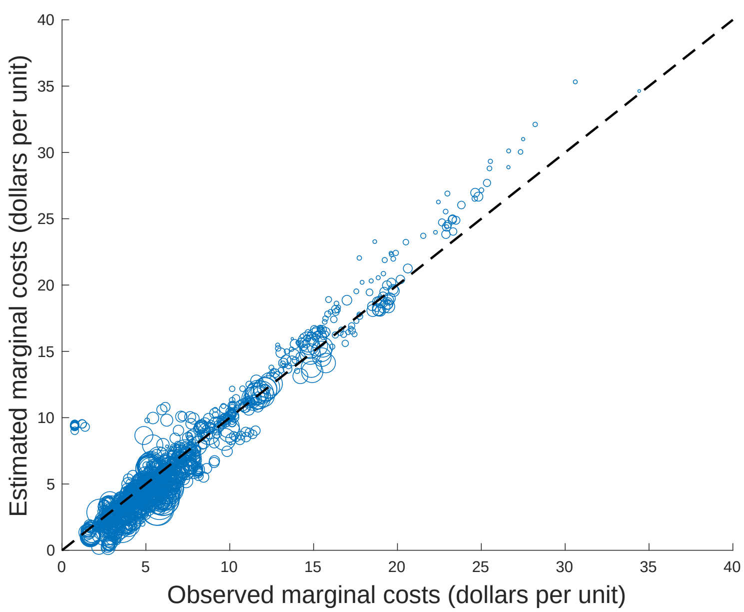
Since we observe wholesale prices at the product level in our dataset, we can compare our estimated marginal costs against the observed values to assess the model's goodness of fit. Figure 1 shows a scatter plot of estimated marginal costs and observed wholesale price. Each data point is represented by a circle of size proportional to the product's market share. The correlation of those points is 0.97. They align almost perfectly with the 45-degree line except for a handful of high-marginal cost products with tiny market shares. Overall, the recovered marginal costs are overestimated by 12.6%. This can be explained by the fact that supermarket marginal costs include other costs (such as labor). We take these goodness-of-fit measures as a strong sign that our depiction of the market structure is highly accurate.

4.4.2 Cross-category correlations

Our methodology does not model shopping behavior across categories. Consumers can have different price sensitivities for different product categories. However, for product categories that are similar, we should expect price sensitivities to be close.

To assess this hypothesis, we first use Bayes's rule to compute each consumer's expected

Figure 1: Estimated vs observed marginal costs



NOTES: Marginal costs observed and estimated from our calibration procedure. Each circle represents one product weighted by its market share.

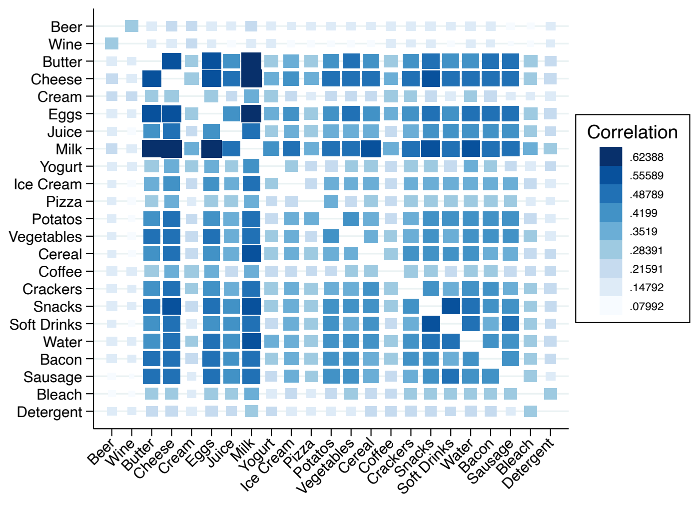
type given their purchase histories in each product category:

$$E(\alpha^k|\mathbf{h}_i) = \sum_{d=1}^{D_k} \alpha_d^k \cdot \phi_{d|\mathbf{h}_i}, \tag{8}$$

where $\phi_{d|\mathbf{h}_i}$ denotes the probability of α_d conditional on history h_i. We then compute the cross-category correlation by looking at the joint distribution of the price sensitivities which can be compared to the correlation in sales observed in the data. To get the most accurate measure of observed cross-category correlation, we look at correlation in observed sales that occur in the same purchase occasion.

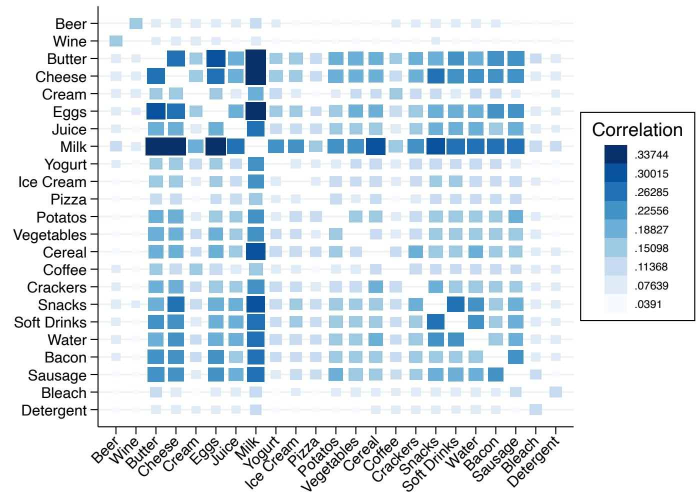
These correlations are shown in Figure 2 and Figure 3. The patterns are similar in both graphs although the correlation is about twice larger when looking at predicted types. Expected consumer types are computed using the full history for each consumer separately. This can lead to overestimating the complementarities between products if for example consumers that purchase two categories often do not purchase them together in the same purchase occasion. Finally, the correlation matrices reproduce some well-known complementarities (milk and cereal, snacks and soda) but fail to reproduce others such as snacks and beer. In general, goods that are storable such as alcohol, coffee, or non-food groceries tend to have a low complementarity with all other goods, while highly perishable goods such as dairy products tend to have a high complementarity with all other goods.

Figure 2: Correlation in $E(\alpha^k|\mathbf{h}_i)$



NOTES: Each square represents the correlation between the vectors of expected consumer types for two given categories.

Figure 3: Correlation in purchases (at same time)



NOTES: Each square represents the correlation between the purchase occasions vectors containing 1s and 0s indicating whether the consumer bought goods from the two product categories at the same time or not.

5 Aggregate effects of personalized pricing

We perform simulations using our calibrated parameters to understand the impact of personalized pricing based on purchase histories. The implicit assumption is that manufacturers do not have access to the consumers' purchase data, only the supermarket has. From the set of first order conditions (4) and for a given history length τ, we can solve for the vector of prices $\mathbf{p}(\mathbf{h}^{\tau})$, which is equivalent to a different price for each different observed purchase history. As described in Section 2, the market shares inside those first order conditions are computed using the posterior densities constructed from the purchase histories. Similarly, we use the set of first order conditions (5) to find prices under perfect discrimination.

In a first set of results, we characterize the information acquisition process of the supermarket as it accumulates more data about consumers behavior. We propose a measure of the firm's learning that helps understanding how close and how fast it can get to perfect price discrimination (in this case third-degree price discrimination). In a second set of results, we explore how personalized pricing distorts the price distribution away from uniform pricing and towards third-degree price discrimination. This allows us to assess the distributional effects that arise from this type of price discrimination. Finally, we look at the aggregate effects on profits, consumer surplus, and welfare.

5.1 Information acquisition

To explore the firm's learning process we construct a simple index that characterizes information acquisition. Our chosen metric is based on the standard deviation of α_d conditional on observing a history of length τ, normalized by the unconditional standard deviation of α_d. We use this normalization for comparability across categories. Formally, our information acquisition metric is

$$\mathcal{I}(\tau) = 1 - \frac{\sigma_{f_{\alpha|\mathbf{h}^\tau}}}{\sigma_{f_\alpha}},$$

where the conditional and unconditional standard deviations are

$$\sigma_{f_{\alpha|\mathbf{h}^\tau}} = \sum_{\ell=1}^{L} \left(\sum_{d=1}^{D_k} \left(\alpha_d - \mathrm{E}(\alpha|\mathbf{h}_\ell^\tau)\right)^2 \cdot \phi_{d|\mathbf{h}_\ell^\tau} \right)^{1/2} \cdot \left(\frac{M_{\mathbf{h}_\ell^\tau}}{M} \right),$$

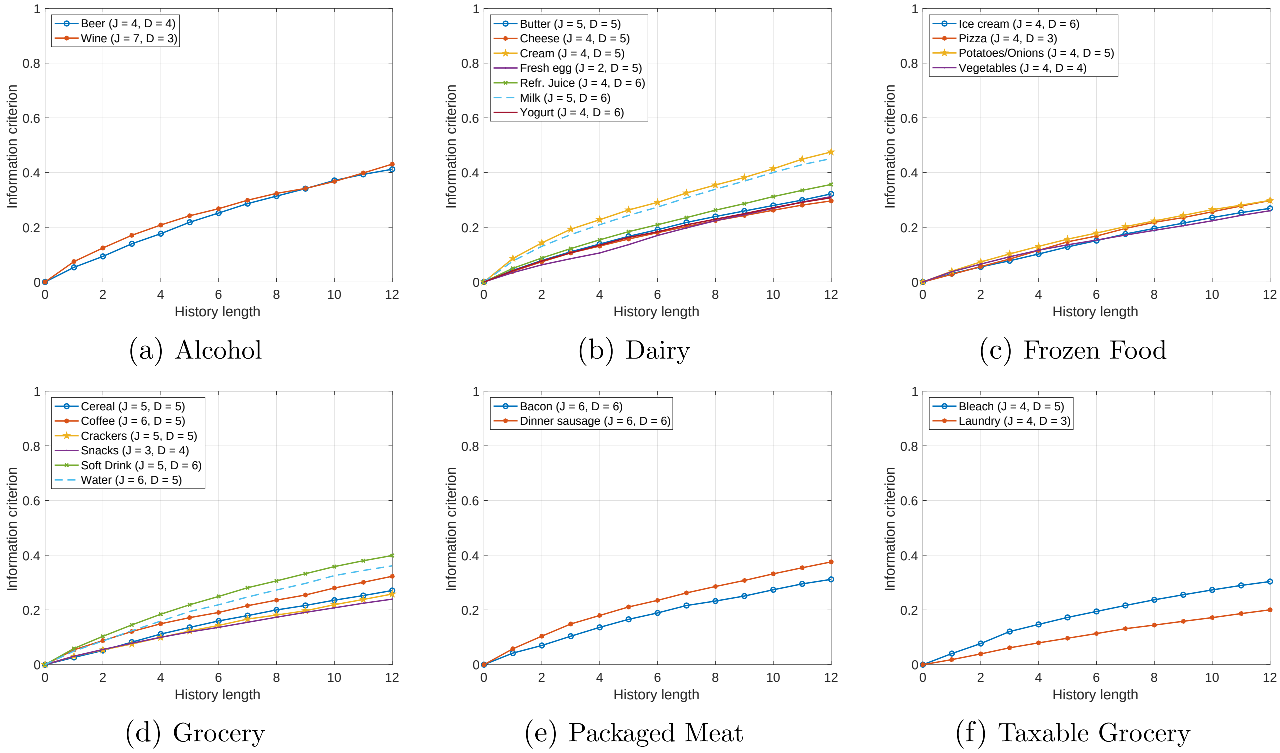
$$\sigma_{f_\alpha} = \left(\sum_{d=1}^{D_k} \left(\alpha_d - \mathrm{E}(\alpha)\right)^2 \cdot \phi_d \right)^{1/2}.$$

This index takes a value of 0 when no information on purchase histories is available and a value of 1 when the firm has perfect information.[10] If information on purchase histories is informative of consumers' types, the information criterion should be increasing in history length and converge towards the perfect price discrimination case.

We now compute our information criterion for each category as a function of the amount of information available to the firm. Results are presented in Figure 4 for a subset of the product categories. A few observations are worth noting. First, histories seem informative of consumer types in the sense that $\mathcal{I}(\tau)$ is strictly increasing in τ for all categories. After observing consumers for 12 periods, the supermarket is able to reduce its uncertainty about consumers by around 30 to 70%, depending on the product category. The second observation is that the returns on information are not necessarily decreasing. For example, information seems to have increasing returns for the "Wine" category over at least part of the range (for $\tau \in [8, 12]$). Finally, we want to highlight that, while the firm seems to be far from identifying consumers perfectly after recording their purchases for one year, information acquisition is asymmetric. The retailer find it easy to assign a type to consumers that purchase something

[10] When the firm has perfect information, $\phi_{d|\mathbf{h}_\ell^\infty}$ becomes degenerate and $\sigma_{f_{\alpha|\mathbf{h}^\infty}} = 0$. Therefore, $\mathcal{I}(\infty) = 1$.

Figure 4: Information acquisition



(a) Alcohol (b) Dairy (c) Frozen Food

(d) Grocery (e) Packaged Meat (f) Taxable Grocery

NOTES: In each panel, the lines represent the information acquisition index for goods in six different product categories as a function of the history length.

in several periods, but struggle to pin down the type of consumers that never purchase anything, which contributes to the large aggregate uncertainty about consumer types.

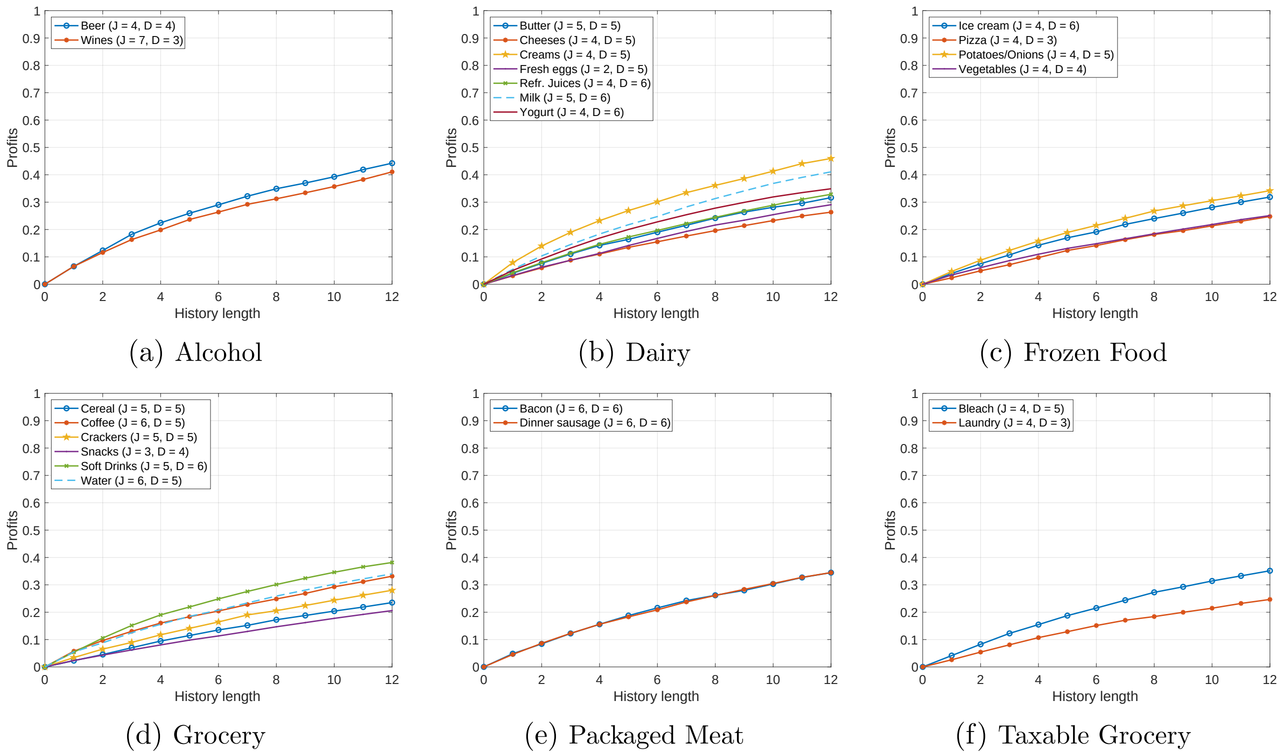
5.2 Impact on profits and consumers

We explore how price discrimination based on purchase histories affects aggregate profits and consumer surplus. To that end, we construct a normalized measure for profits and consumer surplus, namely,

$$\Delta_\pi\left(\mathbf{h}^\tau\right) = \frac{\pi\left(\mathbf{p}^{PS}(\mathbf{h}^\tau)\right) - \pi\left(\mathbf{p}^{UN}\right)}{\left|\pi\left(\mathbf{p}^{PF}\right) - \pi\left(\mathbf{p}^{UN}\right)\right|}$$

$$\Delta_{CS}\left(\mathbf{h}^\tau\right) = \frac{CS\left(\mathbf{p}^{PS}(\mathbf{h}^\tau)\right) - CS\left(\mathbf{p}^{UN}\right)}{\left|CS\left(\mathbf{p}^{PF}\right) - CS\left(\mathbf{p}^{UN}\right)\right|}$$

Figure 5: Impact of targeted pricing on aggregate profits



(a) Alcohol (b) Dairy (c) Frozen Food

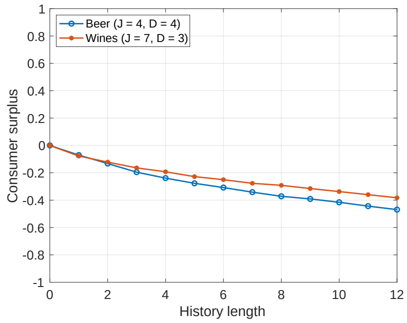
(d) Grocery (e) Packaged Meat (f) Taxable Grocery

NOTES: Each line represents the evolution of the gains in profits as defined by the ratio of the profits from personalized pricing minus profits from uniform pricing divided by the difference between perfect pricing and uniform profits.

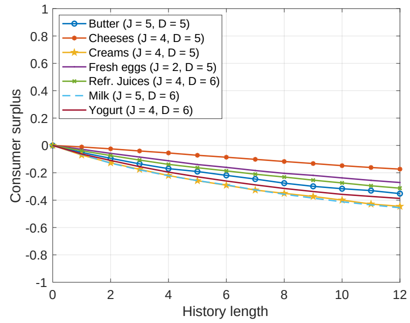
where the absolute value is taken over the denominator to preserve the sign of the numerator, and $\pi(\cdot)$ and $CS(\cdot)$ denote aggregate profits and aggregate consumer surplus.[11] Thus again, our chosen metric allows us to quantify the extent to which profits or consumer surplus converge towards the perfectly discriminatory case. Results on profits from the simulation of personalized price discrimination are presented in Figure 5. It is interesting to see that the supermarket achieves roughly between 60 and 80% of the profits under perfect price discrimination despite there being still significant uncertainty about consumers types. This is a consequence of the asymmetry of the information acquisition process: the retailer is able to guess the type of low-sensitivity agents, and these consumers generate the vast majority of profits. This suggests that the firm may get close to the discriminatory outcome even with limited data.

[11] In all cases, we have that $\pi(\mathbf{p}^{PF}) > \pi(\mathbf{p}^{UN})$. In all cases, we have that $CS(\mathbf{p}^{PF}) < CS(\mathbf{p}^{UN})$.

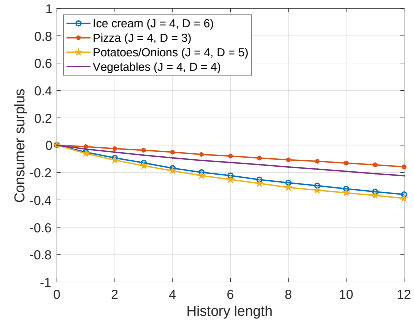
Figure 6: Impact of personalized pricing on consumer surplus



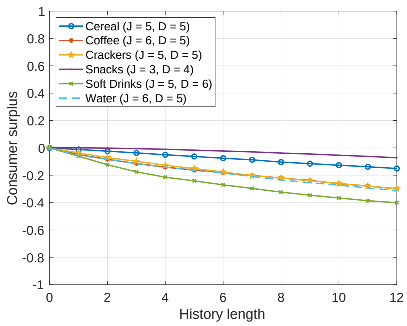
(a) Alcohol



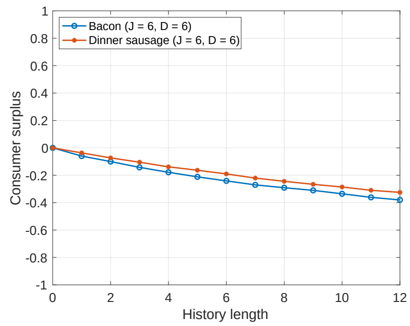
(b) Dairy



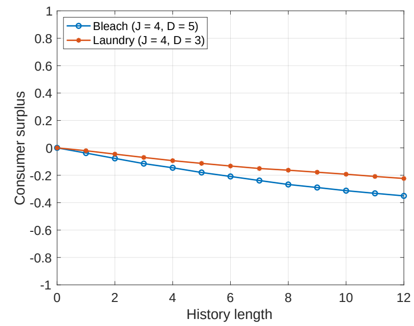
(c) Frozen Food



(d) Grocery



(e) Packaged Meat



(f) Taxable Grocery

NOTES: Each line represents the evolution of the changes in consumer surplus as defined by the ratio of the CS from personalized pricing minus CS from uniform pricing divided by the difference between perfect pricing and uniform pricing CS.

We also consider the impact on consumers in Figure 6. Consumer surplus decreases in all cases except for the *Wine* category, despite that the vast majority of consumers receives a discount over uniform prices. There are two reasons that contribute to this finding. First, the discount offered to high-sensitivity consumers is small in comparison to the premium charged to low-sensitivity consumers. In this context, consumer surplus increases less for the former group than it decreases for the latter. Second, the decrease is amplified also because highly sensitive consumers have a high value of α, and that α is negatively correlated with consumer surplus. Therefore, in addition to receiving a smaller price effect, these consumers' surplus reacts less to changes in price. Rhodes and Zhou (2022) show that under specific circumstances, personalized pricing can increase consumer surplus, if for example it improves market coverage enough. In our case however, the effect of personalized pricing on market coverage is small.

5.3 What drives profitability of personalized pricing?

Table 3: Regression analysis – Learning index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
J_k	0.022***								0.027***
	(0.003)								(0.003)
D_k		-0.003							-0.021***
		(0.003)							(0.005)
$\mathrm{E}(\alpha_d)$			0.038***						-0.059**
			(0.008)						(0.023)
$\mathrm{SD}(\alpha_d)$				-0.200***					-0.322***
				(0.035)					(0.086)
$\mathrm{E}(\bar{\delta}_t)$					0.015***				0.006
					(0.003)				(0.004)
$\mathrm{SD}(\bar{\delta}_t)$						0.002			-0.014***
						(0.004)			(0.005)
HHI							-0.013		0.058**
							(0.025)		(0.024)
Coverage								0.297***	1.199***
								(0.099)	(0.157)
Observations	276	276	276	276	276	276	276	276	276
Within R-squared	.16	.005	.072	.109	.078	.001	.001	.033	.438

NOTE: The dependent variable is the normalized information criterion, $\mathcal{I}(\tau)$, which summarizes how much the retailer learns about consumers from observing purchase histories of length τ. The dataset was assembled by stacking the results from the simulation on 23 categories and 12 history lengths. *HHI* is the normalized Herfindahl-Hirschman Index computed at the product level and *Coverage* is the market share of the inside good. Includes history length fixed effects. Standard error in parenthesis. Significance: * $<$ 0.10; ** $<$ 0.05; *** $<$ 0.01.

In this subsection, we relate key market outcomes to market characteristics in order to understand what drives the profitability of personalized pricing. We achieve this through a descriptive regression analysis. We study four separate outcomes: learning, aggregate profits, aggregate consumer surplus, and the proportion of consumers that receive an infra-marginal offer as a result of personalization.

We correlate these outcomes to market characteristics such as the number of products available, the average product quality, and the degree of product differentiation in each category (as measured by the standard deviation of product quality). Additionally, we also correlate the outcomes to consumer characteristics such as their price sensitivity (both the average and the standard deviation) and the number of consumer types. Finally, we consider

Table 4: Regression analysis – Profitability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9))
J_k	0.024***								0.030***
	(0.006)								(0.005)
D_k		0.039***							0.010
		(0.005)							(0.008)
$\mathrm{E}(\alpha_d)$			0.070***						-0.178***
			(0.014)						(0.037)
$\mathrm{SD}(\alpha_d)$				-0.091					-0.717***
				(0.062)					(0.140)
$\mathrm{E}(\bar{\delta}_t)$					-0.015***				-0.004
					(0.005)				(0.006)
$\mathrm{SD}(\bar{\delta}_t)$						-0.040***			-0.053***
						(0.007)			(0.009)
HHI							0.000		-0.039
							(0.041)		(0.039)
Coverage								1.294***	1.835***
								(0.146)	(0.255)
Observations	276	276	276	276	276	276	276	276	276
Within R-squared	.068	.224	.091	.008	.031	.113	0	.229	.457

NOTE: The dependent variable is the normalized measure of profits, $\Delta_\pi(\tau)$, which summarizes how close to the perfect discrimination case the retailer gets from observing purchase histories of length τ. The dataset was assembled by stacking the results from the simulation on 23 categories and 12 history lengths. *HHI* is the normalized Herfindahl-Hirschman Index computed at the product and *Coverage* is the market share of the inside good. Includes history length fixed effects. Standard error in parenthesis. Significance: * $<$ 0.10; ** $<$ 0.05; *** $<$ 0.01.

the Herfindahl-Hirschman Index[12] (henceforth HHI) and market coverage[13] in the set of controls. In our context the HHI does not measure concentration in the traditional sense since the retailer is a monopolist. Instead, we compute it at the product level to capture the extend to which one or a few products dominate the market over others. Results are presented in Table 3 to Table 6. We report coefficient estimates and the within R-squared, which allows us to report on the explanatory power of each market characteristic.

Learning. We first study the determinants of learning in Table 3. We find that product variety and the spread in consumer types are the main drivers of learning. One more product available is associated with an increase of our learning index of 0.03, and product variety explains 16% of the variation in learning. The standard deviation of consumers' price sensitivity is the second most important driver of learning and impacts the outcome negatively.

[12] We use the normalized Herfindahl-Hirschman Index,

$$HHI = \frac{\sum_{j=1}^{J} \left(\frac{s_j}{1-s_0} \right)^2 - \frac{1}{J}}{1 - \frac{1}{J}},$$

which takes values over the interval $[0, 1]$ for comparability between categories.

[13] Defined as the market share of the inside good.

Table 5: Regression analysis – Aggregate consumer surplus

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
J_k	0.020*								0.005
	(0.011)								(0.007)
D_k		-0.118***							-0.092***
		(0.006)							(0.013)
E(α_d)			-0.036						0.380***
			(0.027)						(0.056)
SD(α_d)				-0.454***					1.295***
				(0.113)					(0.211)
E($\overline{\delta}_t$)					0.091***				0.034***
					(0.008)				(0.009)
SD($\overline{\delta}_t$)						0.086***			0.074***
						(0.013)			(0.013)
HHI							-0.232***		0.122**
							(0.076)		(0.059)
Coverage								-3.010***	-1.939***
								(0.254)	(0.384)
Observations	276	276	276	276	276	276	276	276	276
Within R-squared	.013	.57	.007	.057	.31	.143	.034	.348	.655

NOTE: The dependent variable is the normalized measure of consumer surplus, $\Delta_{CS}(\tau)$, which summarizes how close to the perfect discrimination case the retailer gets from observing purchase histories of length τ. The dataset was assembled by stacking the results from the simulation on 23 categories and 12 history lengths. *HHI* is the normalized Herfindahl-Hirschman Index computed at the product and *Coverage* is the market share of the inside good. Includes history length fixed effects. Standard error in parenthesis. Significance: * $<$ 0.10; ** $<$ 0.05; *** $<$ 0.01.

Market coverage positively impacts learning, however its explanatory power is small (3.3%). The effect of other market characteristics is either small or insignificant in terms of explanatory power. Taken altogether, the market characteristics explain 43.8% of the supermarket's learning defined by our information metric.

Profitability. We next move to profitability as defined in equation (2). These results are found in Table 4. We find that market coverage is the main driver of profits, followed by the number of consumer types, product differentiation, and the average price sensitivity of consumers. The supermarket can extract more profits from consumers when there are more consumer types and when consumers are on average less price sensitive. Also, the monopolist can increase profits more when the available products are more homogeneous (i.e. less differentiated). These findings are consistent with economic theory. Taken altogether, the market characteristics explain 45.7% of the variation in the profitability of personalized pricing.

Consumer surplus. The results are similar for aggregate consumer surplus, Table 5, which makes sense since the firm increases profitability to the detriment of consumers when using discriminatory pricing. We find opposite effects than for profits, namely, personalization

Table 6: Regression analysis – Share of consumers with personalized prices below uniform prices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
J_k	0.059***								0.058***
	(0.009)								(0.009)
D_k		-0.044***							-0.067***
		(0.008)							(0.015)
E(α_d)			0.083***						0.233***
			(0.023)						(0.067)
SD(α_d)				-0.515***					0.621**
				(0.096)					(0.252)
E($\overline{\delta}_t$)					0.052***				0.023**
					(0.008)				(0.010)
SD($\overline{\delta}_t$)						0.007			0.004
						(0.012)			(0.016)
HHI							-0.146**		0.210***
							(0.066)		(0.070)
Coverage								-0.824***	0.148
								(0.268)	(0.458)
Observations	276	276	276	276	276	276	276	276	276
Within R-squared	.149	.103	.048	.098	.132	.001	.018	.035	.35

NOTE: The dependent variable is the share of consumers that receive infra-uniform prices under personalized pricing and for which consumer surplus increase, given a purchase history of length τ. The dataset was assembled by stacking the results from the simulation on 23 categories and 12 history lengths. *HHI* is the normalized Herfindahl-Hirschman Index computed at the product and *Coverage* is the market share of the inside good. Includes history length fixed effects. Standard error in parenthesis. Significance: * < 0.10; ** < 0.05; *** < 0.01.

harms consumers less when they are more homogeneous and products are more diversified. Market coverage is also a very important driver of consumer surplus. In particular, we find that consumer surplus is higher when market coverage is low and there is room for a larger market expansion effect. Taken altogether, our market characteristics explain 65.5% of the variation in consumer surplus from personalized prices.

Infra-uniform price consumers. We turn our attention to our final market outcome, the share of consumers that receive an infra-uniform offer, Table 6. This is important in terms of equity: our analysis suggests that most consumers receive a small rebate while a small subset of price insensitive consumers are charged very high premia on prices. We find that the number of products available, the number of consumer types and their dispersion, and average product quality are the biggest drivers in this case. Taken altogether, our market characteristics explain only around 35% of the variation in the share of infra-uniform consumers.

6 Price dispersion and heterogenous effects across consumers

A main concern with price discrimination practices is the potential harm to consumers relative to the uniform pricing environment. In particular, one might be concerned about large differences in consumer surplus extraction across consumers because those differences might be correlated with income or certain demographic characteristics.

We start this section by providing bounds on the price response of the retailer using our calibrated model, and document the degree of asymmetry in prices that arise from price discrimination. We also provide graphical evidence on how fast prices converge to the perfect discrimination case. This allows us to highlight another asymmetry in the information acquisition process: the firm learns faster from price insensitive consumers that purchase often relative to price sensitive consumers that almost never shop. In other words, information about the consumer's type is revealed when a purchase occurs.

Although we do not have access to detailed information on consumers' characteristics, we use our model to study the impact of price discrimination at the consumer level. We investigate further how price discrimination affects consumers total purchases and consumer surplus, and how much profits the retailer can extract per consumer. We find evidence of a large amount of heterogeneity at the consumer level.

6.1 Price dispersion

We first describe the distribution of prices provided that the retailer observed one full year of purchases. The results are presented in Table 7. We report the average uniform price per category (in dollars), the maximum discount offered to the most price sensitive consumers, and the maximum premium charged to price insensitive consumers (in %). Finally, we report the share of consumers that received a discount.

We find that the distribution of prices is highly asymmetric in all product categories under study. The overwhelming majority of consumers receive a discount when the firm is allowed to price discriminate using their purchase histories. However, the discounts offered are small (8.6% on average). On the other hand, price sensitive consumers pay a large premium over uniform prices. The maximum surcharge is 29.9% on average, roughly 3 times as large as the maximum discount. The retailer thus extracts a large surplus from a small subset of consumers she identifies as being price insensitive based on their purchase patterns.

Table 7: Price dispersion

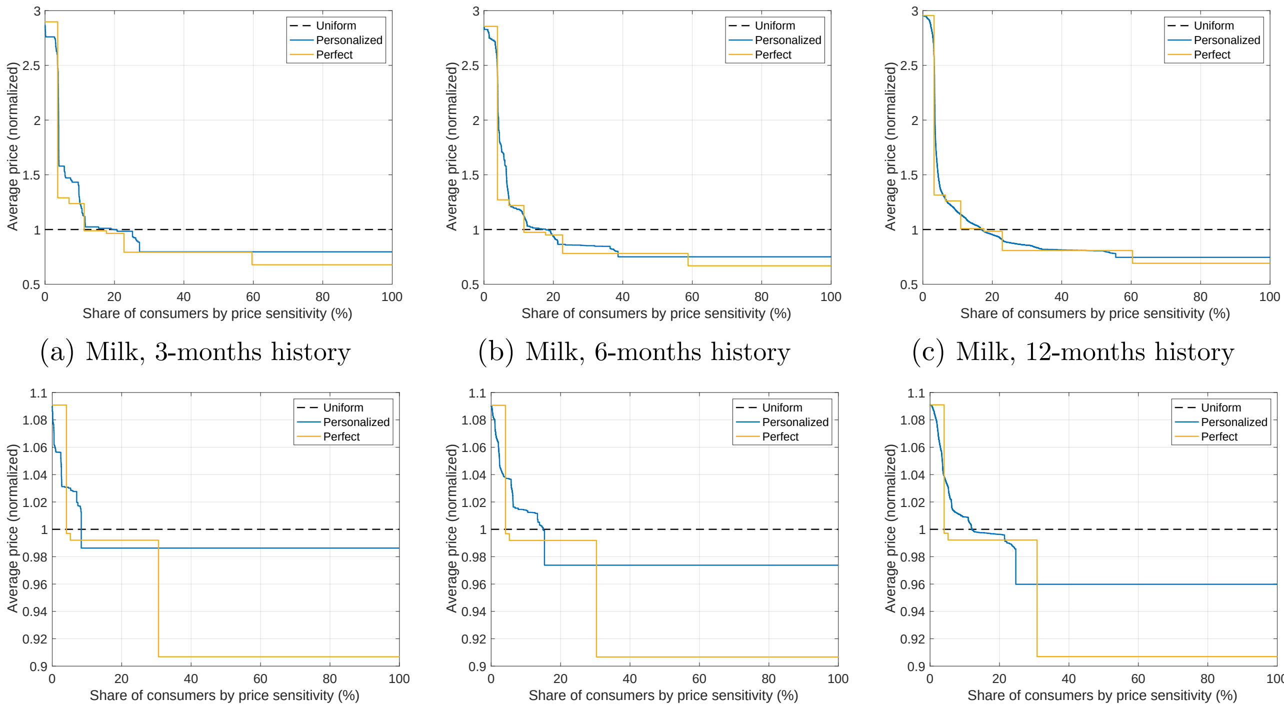
Category	Uniform price	Max. discount (%)	Max. surcharge (%)	Share infra-uniform
Alcohol				
Beer	21.24	4.0	9.0	0.97
Wines	25.44	6.7	5.3	0.94
Dairy				
Butter/Margarine/Spreads	4.90	10.1	35.3	0.90
Cheese	7.22	15.1	50.5	0.80
Creams/Creamers	5.90	7.3	19.8	0.88
Fresh eggs	5.05	11.2	41.5	0.85
Refrigerated Juice/Beverage	7.21	12.1	39.7	0.88
Milk	8.20	25.4	195.4	0.83
Yogurt	6.78	5.6	15.5	0.90
Frozen food				
Ice Cream	7.71	7.7	30.6	0.90
Frozen Pizza	9.20	4.8	14.3	0.90
Frozen Potatoes/Onions	5.38	7.9	18.7	0.94
Frozen Vegetables	9.93	12.1	21.8	0.89
Grocery				
Cereal	6.00	4.0	9.1	0.88
Coffee	11.11	3.8	8.2	0.96
Crackers	4.56	5.0	15.6	0.90
Snacks grocery	8.52	6.2	10.4	0.79
Soft drinks/Mixers	10.75	9.7	35.3	0.80
Water	7.66	8.9	18.9	0.88
Packaged meat				
Bacon	10.24	7.1	21.9	0.93
Dinner sausage	11.79	16.5	61.8	0.89
Taxable grocery				
Bleach/Stain removers	4.70	4.6	7.7	0.96
Laundry detergent	9.64	1.6	2.5	0.93

6.2 Price dispersion and information acquisition

Next, we want to understand the effect that price discrimination has on consumers as a function of their predicted type and how pricing changes as there is more information acquisition. We focus on two representative categories, "Milk" and "Cereal", to provide some intuition about the pricing mechanism. Results are presented in Figure 7. These figures represent the distribution of prices over consumers, ranked by price sensitivity, after observing a purchase history of three, six, and twelve months respectively. All graphs present the average personalized price, the average perfectly discriminatory price, and the average uniform price.[14] All prices are normalized by dividing by the average uniform price for comparability across categories.

[14]All prices are averaged over the J_k products with equal weights.

Figure 7: Price dispersion, by history length



(a) Milk, 3-months history

(b) Milk, 6-months history

(c) Milk, 12-months history

(d) Cereal, 3-months history

(e) Cereal, 6-months history

(f) Cereal, 12-months history

NOTES: Distributions of prices over consumers, ranked by price sensitivity, after observing a purchase history of three, six, and twelve months respectively.

There are a few things to note. First, the vast majority of consumers receive a personalized price that is below the uniform price. Across all categories, this number varies between 80 and 95% of consumers. Second, the discount offered to these consumers is small in comparison to the premium charged to price insensitive consumers. If we look at “Milk” for example, price insensitive consumers can pay up to 50% more for the same product, while price sensitive consumers receive a discount in the order of 10%. This level of asymmetry is observable in other categories. Third, while prices seem to converge relatively quickly to the perfect discrimination prices for consumers that are charged a premium, the same is not true for consumers receiving a discount. In almost every category, prices do not reach the lowest level suggested by third-degree price discrimination. The reason for this phenomenon is that the supermarket cannot distinguish very well types with high price sensitivity from each other, since these consumers never purchase anything. In this case, the small uncertainty about types leads the supermarket to charge a higher price than what is optimal under perfect discrimination for a large set of consumers.

6.3 Heterogeneity of effects on consumers

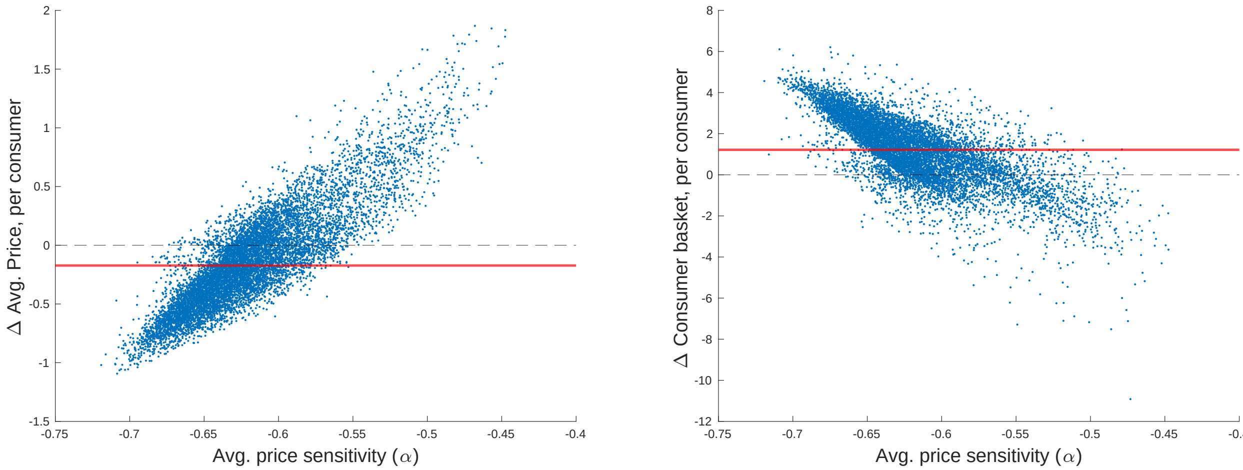
To study how personalized pricing affects consumers at the individual level over their full consumer basket, we assign each consumer an average price sensitivity calculated over categories as follows,

$$\overline{\alpha}_i = \sum_{k=1}^{K} \mathrm{E}(\alpha^k|\mathbf{h}_i) \cdot w_k,$$

where the conditional expectation is calculated as in equation (8), and w_k are category weights based on the category revenues observed in the data. We then plot these price sensitivities against the average price premium/discount offered to these consumers and separately against the average change in consumer expenditures (i.e. the value of each consumer's basket, in dollars). The results are presented in Figure 8. We observe that, consistent with intuition, consumers with a high price sensitivity receive a price discount, which leads them to increase monthly expenditures. This suggests that the substitution effect (from the outside good) is larger than the income effect (from the decrease in price). Consumers with a low price sensitivity exhibit the reverse pattern, but the magnitude is larger. The price effect can be as large as 3.5 times larger for the group that pays a premium relative to the group that gets a discount. Interestingly, our graph suggests that a group of consumers both receives very high prices and increases their monthly spendings. These are located in a separate cloud of points above the main group of consumers. For some of these consumers, the income effect dominates the substitution effect: prices and total expenditure increase.

Finally, we look at how personalized prices affect surplus at the individual level. Results are presented in Figure 9. Panel (a) depicts how much surplus is extracted from each consumer and transformed into profits by the retailer, while panel (b) presents the change in consumer surplus. We find that the supermarket is able to extract at least some surplus from each and every consumer. The average is around 40 cents per consumer (the average monthly expenditure is $12.20 per consumer). However, there is a lot of variation: the retailer is able to extract as much as $10 in extra profits from the most price insensitive consumers. Consumer surplus on the other hand slightly increases for price-sensitive consumers. In this case, price discrimination leads to a win-win scenario. For price-insensitive consumers, consumer surplus decreases and the magnitude is around twice as large as the gains in profits.

Figure 8: Effect of personalized pricing on prices and expenditures, by price sensitivity

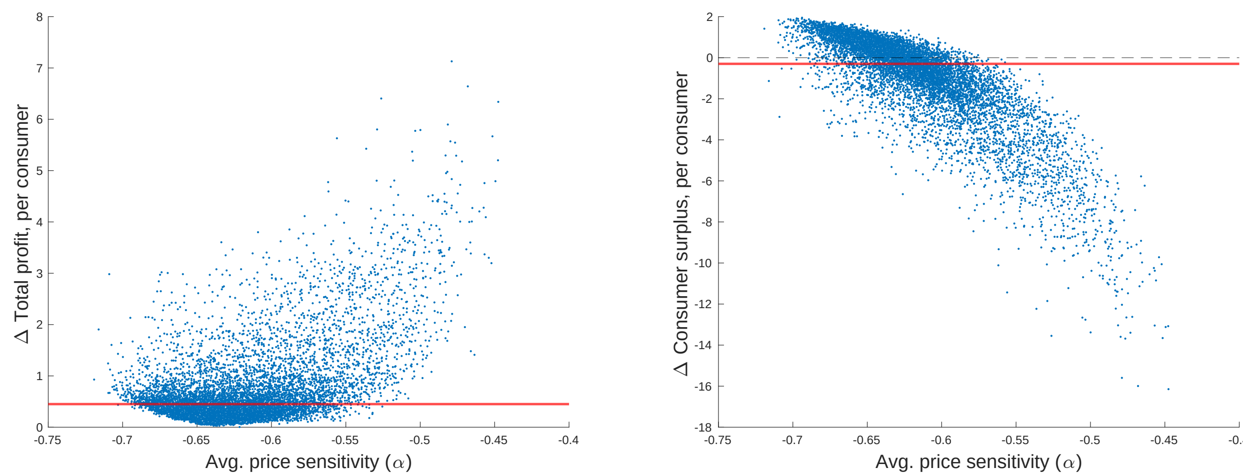


(a) Average price

(b) Monthly consumer expenditure

NOTES: These figures depict the change in average prices and monthly consumer expenditure that occurs when the supermarket switches from uniform pricing to personalized pricing. Each marker represents a unique consumer in our database, ranked by their expected price sensitivity over all categories. The red line represents the average over all consumers. All results are computed assuming a span of 12 months of consumer purchase histories.

Figure 9: Effect of personalized pricing on profits and consumer surplus, by price sensitivity



(a) Profits extracted (per consumer)

(b) Consumer surplus

NOTES: These figures depict the change in per-customer profits and consumer surplus that occurs when the supermarket switches from uniform pricing to personalized pricing. Each marker represents a unique consumer in our database, ranked by their expected price sensitivity over all categories. The red line represents the average over all consumers. All results are computed assuming a span of 12 months of consumer purchase histories.

7 Conclusion

We quantify the dispersion of consumers' price sensitivity across a large number of product categories from a supermarket chain. With the consumer types fully characterized in hand, we implement personalized prices conditional on those consumer types by maximizing profits for each product category and using Bayes' rule to use the purchase histories data from each new time period. We show that once the supermarket introduces personalized prices, it can reap between 60 and 80% of the profits that it would obtain under perfect price discrimination. Those gains in profits originate from the highly price-insensitive consumers who get exposed to supra-uniform prices. However, a large share of consumers are presented with infra-uniform prices.

References

Ali, S. N., Lewis, G., and Vasserman, S. (2022). Voluntary disclosure and personalized pricing. *The Review of Economic Studies*, 90(2):538–571.

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

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 60(4):889–917.

Chandra, A. and Lederman, M. (2018). Revisiting the relationship between competition and price discrimination. *American Economic Journal: Microeconomics*, 10(2):190–224.

Chen, Z., Choe, C., and Matsushima, N. (2020). Competitive personalized pricing. *Management Science*.

Choe, C., Matsushima, N., and Tremblay, M. J. (2022). Behavior-based personalized pricing: When firms can share customer information. *International Journal of Industrial Organization*, 82:102846.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics*, 13(2):545–591.

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

Döpper, H., MacKay, A., Miller, N., and Stiebale, J. (2023). Rising markups and the role of consumer preferences.

Dubé, J.-P., Fang, Z., Fong, N., and Luo, X. (2017). Competitive price targeting with smartphone coupons. *Marketing Science*, 36(6):944–975.

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

Dubé, J.-P. and Misra, S. (2023). Personalized pricing and consumer welfare. *Journal of Political Economy*, 131(1):131–189.

Eizenberg, A., Lach, S., and Oren-Yiftach, M. (2021). Retail prices in a city. *American Economic Journal: Economic Policy*, 13(2):175–206.

Ershov, D., Laliberté, J.-W., Marcoux, M., and Orr, S. (2021). Estimating complementarity with large choice sets: An application to mergers. *Mathieu and Orr, Scott, Estimating Complementarity With Large Choice Sets: An Application to Mergers (March 10, 2021).*

Fudenberg, D. and Villas-Boas, J. M. (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436.

Goolsbee, A. and Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica*, 72(2):351–381.

Grieco, P. L., Murry, C., Pinkse, J., and Sagl, S. (2022). Conformant and efficient estimation of discrete choice demand models. *Penn State*.

Hitsch, G. J., Hortacsu, A., and Lin, X. (2019). Prices and promotions in us retail markets: Evidence from big data. Technical report, National Bureau of Economic Research.

Jullien, B., Reisinger, M., and Rey, P. (2023). Personalized pricing and distribution strategies. *Management Science*, 69(3):1687–1702.

Miller, N. H., Remer, M., Ryan, C., and Sheu, G. (2016). Pass-through and the prediction of merger price effects. *The Journal of Industrial Economics*, 64(4):683–709.

Miller, N. H., Remer, M., and Sheu, G. (2013). Using cost pass-through to calibrate demand. *Economics Letters*, 118(3):451–454.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2):307–342.

Puller, S. L. and Taylor, L. M. (2012). Price discrimination by day-of-week of purchase: Evidence from the us airline industry. *Journal of Economic Behavior & Organization*, 84(3):801–812.

Rhodes, A. and Zhou, J. (2022). Personalized pricing and competition. *Available at SSRN 4103763*.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340.

Shiller, B. R. (2020). Approximating purchase propensities and reservation prices from broad consumer tracking. *International Economic Review*, 61(2):847–870.

Shin, S., Misra, S., and Horsky, D. (2012). Disentangling preferences and learning in brand choice models. *Marketing Science*, 31(1):115–137.

Smith, A. N., Seiler, S., and Aggarwal, I. (2022). Optimal price targeting. *Marketing Science*.

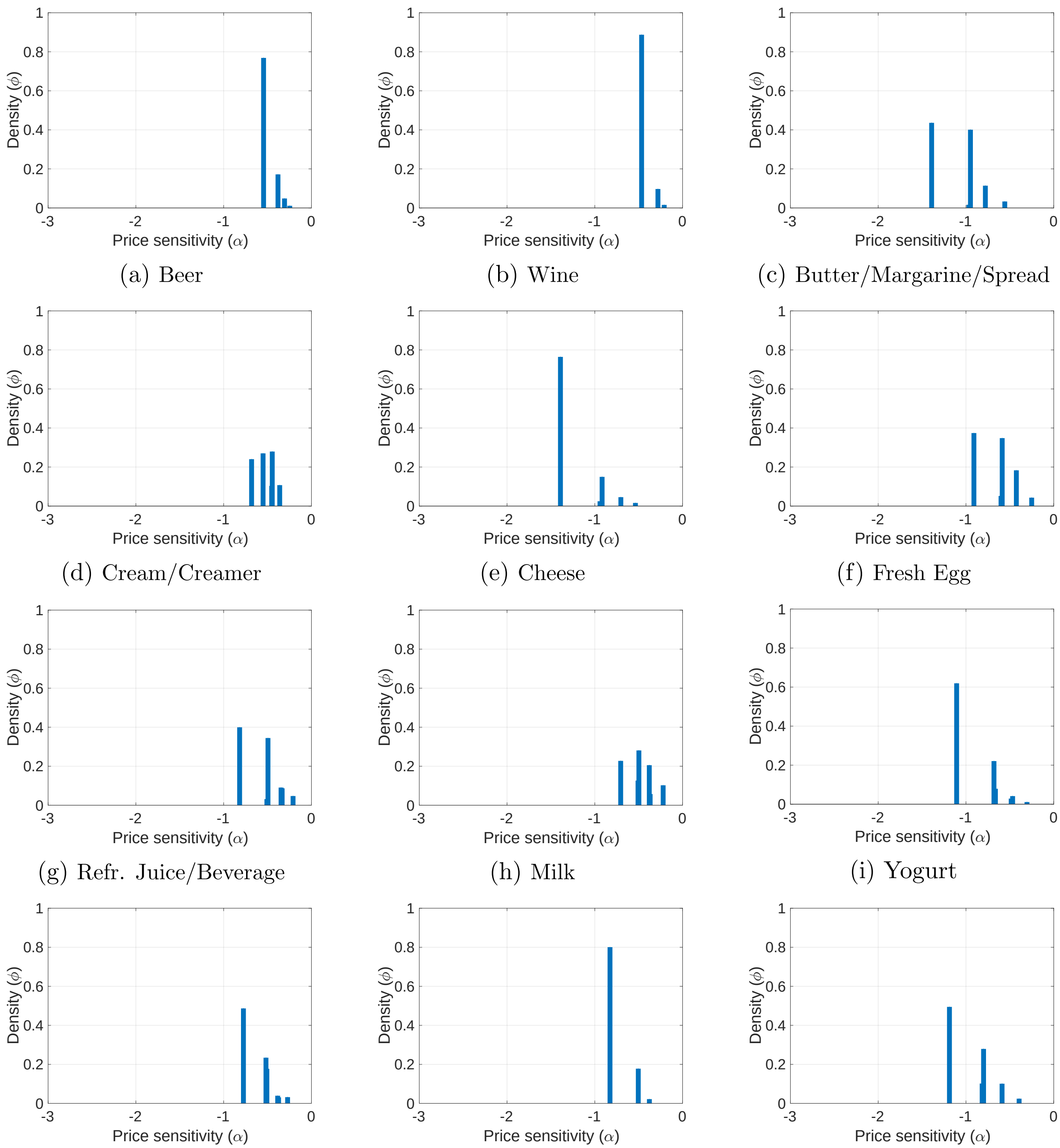
Thomassen, Ø., Smith, H., Seiler, S., and Schiraldi, P. (2017). Multi-category competition and market power: a model of supermarket pricing. *American Economic Review*, 107(8):2308–51.

Waldfogel, J. (2015). First degree price discrimination goes to school. *The Journal of Industrial Economics*, 63(4):569–597.

Appendices

A Additional figures

Figure A.1: Calibrated distributions of price sensitivities



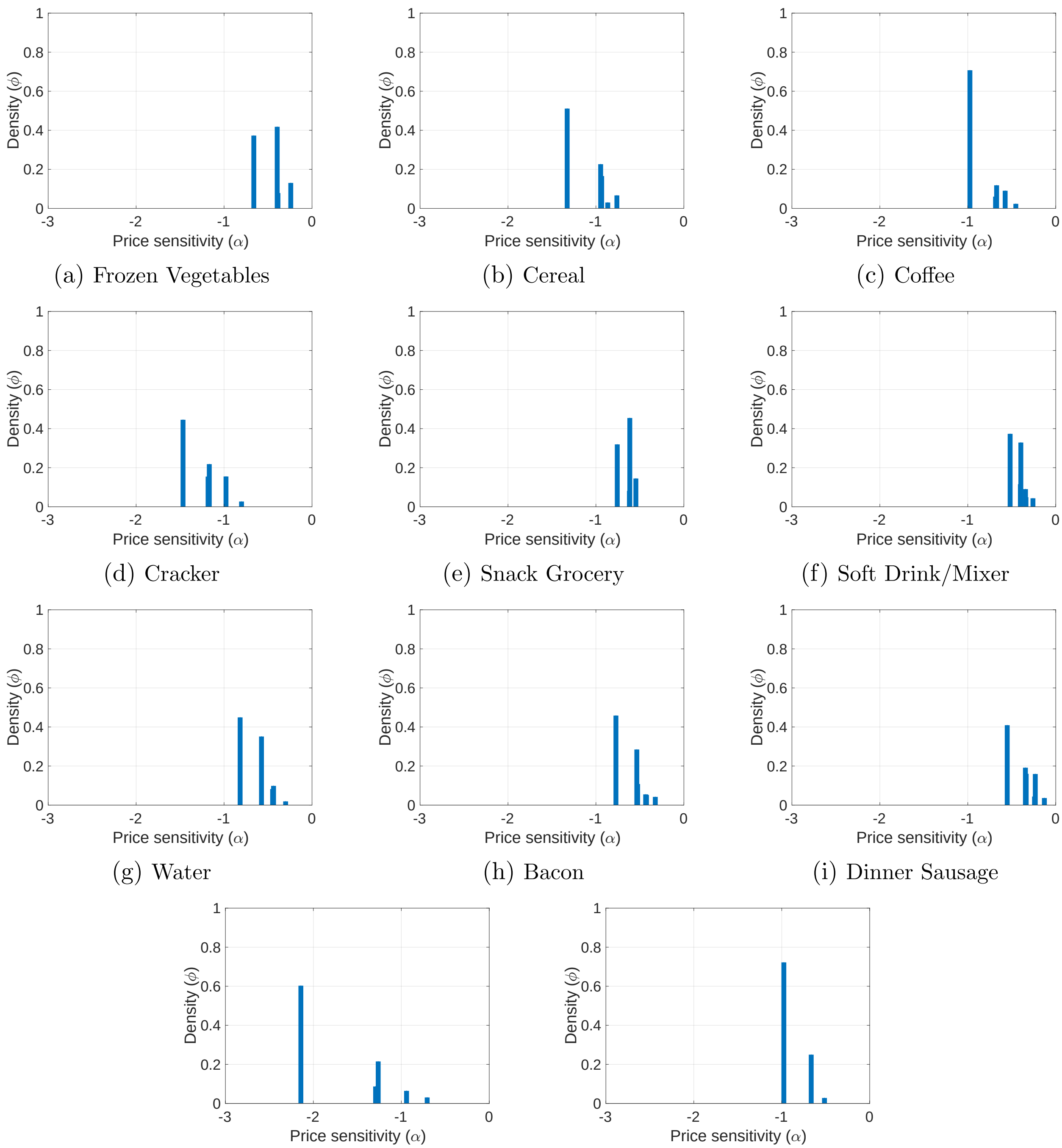
(a) Beer (b) Wine (c) Butter/Margarine/Spread

(d) Cream/Creamer (e) Cheese (f) Fresh Egg

(g) Refr. Juice/Beverage (h) Milk (i) Yogurt

(j) Ice Cream (k) Frozen Pizza (l) Frozen Potatoes/Onions

Figure A.2: Calibrated distributions of price sensitivities



(a) Frozen Vegetables (b) Cereal (c) Coffee (d) Cracker (e) Snack Grocery (f) Soft Drink/Mixer (g) Water (h) Bacon (i) Dinner Sausage (j) Bleach/Stain Remover (k) Laundry Detergent

B Details on the data

B.1 Data

The data used for this analysis come from DecaData. They contain daily point-of-sale transactions from 853 stores in 13 U.S. states (the "ODYSSEY" database), retailer product deliveries and restocking at these stores (the "ARGO" database), and detailed product information (the "PRODUCT" database). All data are recorded daily at the universal product code (henceforth UPC).

ODYSSEY database. The ODYSSEY database describes all purchases realized at the 853 stores. We focus on the year 2018. The data include the date of transaction, a store identifier, the UPC, quantity purchased, price, as well as a transaction identifier that allows to identify joint purchases. Additionally, we observe a consumer identifier for a subset of consumers who are enrolled in a loyalty card program. We use these consumer identifiers to recover consumers purchase histories.

ARGO database. The ARGO database contains wholesale purchases from the 853 stores and include the date of delivery, the store identifier, the UPC, quantity ordered and delivered, as well as the unit wholesale price of the goods delivered. The database spans from December 2017 and all of 2018.

We use the ARGO database to recover wholesale prices for the products available in the ODYSSEY database. This is done as follows. We begin by aggregating the ARGO database at the UPC-store-month level to recover the weighted average wholesale price of each product in each month for each store. We believe that these monthly averages capture reasonably well the wholesale prices faced by each store. In some cases, no restocking of a product occurs in the same month a sale is recorded. When this happens, we match the wholesale price from the previous month instead.

PRODUCT database. Finally, we have access to a database containing detailed information on the 400,000 products present in the ODYSSEY and ARGO databases. Products are defined at the UPC level and are grouped into 56 departments (e.g, dairy), 454 categories (e.g., milk), and 2,646 subcategories (e.g., organic milk). We also observe the manufacturer and the brand names as well as the package size and units of measurement (e.g, 16 oz).

One limitation with our dataset is that it does not include product characteristics (e.g. the nutritional content, the amount of sugar, calories). Observed products characteristics have been used to construct instruments to identify the price coefficient in demand estimation since Berry et al. (1995).

B.2 Sample selection and summary statistics

We focus on one representative supermarket to perform the analysis. We choose a store in the vicinity of the median store in terms of total sales. Our chosen store had $13.1 million in sales in 2018, placing it at the 52$^{\text{nd}}$ percentile of the distribution. We also restrict our attention to the 17,756 consumers enrolled in a loyalty program and for whom we can track purchases over time.[15] These consumers account for about 79% of all revenues recorded at our representative store in 2018.

We further restrict our sample as follows. We focus our attention on 6 of the 56 food departments. The chosen departments are alcohol, dairy products, frozen food, grocery, packaged meat, and taxable (non-food) grocery. The choice of departments is guided by the match quality between the sales and the costs data, and we drop departments for which we could not recover the wholesale prices for more than 20% of products. Wholesale prices are not used in the analysis explicitly, however they are useful to assess the fit of our calibrated parameters by comparing observed and predicted wholesale prices. We then restrict our attention to the largest food categories within the chosen food departments, based on sales. Categories that account for less than 5% of total sales (in dollars) by food department are removed.[16] Our final dataset includes 23 food categories.

To restrict the number of products available in each category, we aggregate products over UPCs at the manufacturer level within each category. We then assume that each manufacturer sells a differentiated version of the good as defined by the product category, and we group all manufacturers with less than 5% of total category sales into a single fringe manufacturer. The typical category features between 2 and 7 differentiated products, with an average of 4.6. Since we aggregate over products in different package sizes, we normalize the quantities purchased, unit prices and wholesale prices in terms of the most common package size in each product category.

[15] We eliminated a small number of customer IDs with unreasonably high monthly purchases. We assume these to be institutional clients such as restaurants or other types of firms.

[16] The threshold is set to 2.5% for grocery and taxable non-food groceries.

Table B.1: Per consumer spending for representative store

	Full Dataset	Selected Departments	Selected Categories	Final Dataset
Consumer basket				
– Mean	30.17	18.18	7.77	7.56
– Median	17.29	9.58	3.28	3.00
Monthly spending, conditional on purchase				
– Mean	105.86	63.76	27.28	26.55
– Median	62.14	34.36	14.66	14.16
Monthly spending, unconditional				
– Mean	48.66	29.31	12.54	12.20
– Median	0	0	0	0
Nb. of monthly visits				
– Mean	1.52			
– Median	0			
Total sales	10.4M	6.27M	2.68M	2.61M
Nb. of consumers	17,756			
Store ID	1697			

NOTES: All values reported are computed for consumers enrolled in a loyalty card program in 2018. *Consumer basket* is the total consumer expenditure on a single shopping occasion. *Monthly spending* regroup all purchases made by a given consumer in a given month, and is reported both conditional on visiting the store at least once in the month, or unconditionally. Column 2 restricts the data to the 6 food departments under study. Column 3 restricts the data to the 23 food categories under study within the chosen departments. Column 4 represents our final dataset.

Table B.1 presents some consumer- and store-level summary statistics for the initial dataset and at each step of our selection procedure. The average and median consumer basket on a single shopping occasion are \$30.17 and \$17.29 respectively. Consumers visit the representative store on average 1.52 times per month. Meanwhile, our analysis is conducted by assuming consumers get a single shopping occasion per month. We report monthly consumer spending, both conditionally on visiting the store at least once in a given month and unconditionally. Consumers spend on average \$105.89 per month conditional on a visit (median is \$62.14) and \$48.66 unconditionally (median is \$0). Across all metrics, we see that our final dataset captures roughly one quarter of all purchases made by registered consumers, or one quarter of the average consumer basket.

C Heterogenous shopping probabilities

In the model presented in section 2, we impose the underlying assumption that consumers have a homogeneous probability to visit the shop in a given period, that is,

$$\Pr(\text{consumer } i \text{ visit the store in period } t) = \rho_i = \rho, \quad \forall i = 1, ..., M, \quad \forall t = 1, ..., T.$$

In this section, we delve deeper to understand which forms of consumer heterogeneity allow for this simplified representation, and when this assumption is unreasonable.

We begin by introducing some new notation. Consider the random variable $\omega_i \in \Omega$ and define the event of visiting the shop as

$$\mathbb{1}_S(\omega_i) = \begin{cases} 1 \text{ (shop)} & \text{if } \omega_i \geq S \\ 0 \text{ (not shop)} & \text{otherwise.} \end{cases}$$

We can compute the probability to visit the store as

$$\Pr(\omega \geq S) = \int_\Omega \mathbb{1}_S(\omega_i) dF_\omega(\omega) = \rho_i.$$

In what follows, we consider consumer types to be an (ω, α)-tuple, that is, a profile of propensity to visit the shop and price sensitivity. Recall that the choice probability conditional on visiting the store is

$$s_j(\alpha_i) = \frac{\exp(\delta_j + \alpha_i p_j)}{1 + \sum_{j'} \exp(\delta_{j'} + \alpha_i p_{j'})},$$

and aggregating over $s_j(\alpha_i)$ yields the following market share for product j,

$$s_j = \int_{\underline{\alpha}}^{\overline{\alpha}} \int_\Omega \mathbb{1}_S(\omega_i) s_j(\alpha_i) dF_{\omega,\alpha}(\omega_i, \alpha_i).$$

We fix ideas, we consider three simple cases. First, we consider the case where $\rho_i = \rho$ for all $i = 1, ..., M$, which is equivalent to the case where the random variable ω_i has the same distribution for all i. Notice that, as a consequence, the random variables ω_i and α_i are automatically orthogonal (and uncorrelated). Second, we relax the assumption that $\rho_i = \rho$, but we maintain the orthogonality between ω_i and α_i. One way to interpret this orthogonality condition is that the probability to visit the store doesn't depend on the price sensitivity and vice versa. Finally, we consider a case where ω_i and α_i are not orthogonal. Without loss of generality, we introduce dependence in the following way: the probability to visit the shop is $\rho_i = \rho(\alpha_i)$. We consider each case in turn.

CASE 1: Homogeneous shopping probabilities. In this case, we have the following representation of the market share,

$$\begin{aligned}
s_j &= \int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\Omega} \mathbb{1}_S(\omega_i) s_j(\alpha_i) dF_{\omega,\alpha}(\omega_i, \alpha_i) \\
&= \int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\Omega} \mathbb{1}_S(\omega) s_j(\alpha_i) dF_{\omega}(\omega) dF_{\alpha}(\alpha_i) \\
&= \left(\int_{\Omega} \mathbb{1}_S(\omega) dF_{\omega}(\omega) \right) \cdot \left(\int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha_i) dF_{\alpha}(\alpha_i) \right) \\
&= \rho \cdot \left(\int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha_i) dF_{\alpha}(\alpha_i) \right)
\end{aligned}$$

which is exactly the representation we use in section 2.

We want to address a potential selection issue related to estimating ρ consistently. We consider the following estimator for ρ,

$$\begin{aligned}
\widetilde{\rho} &= \frac{1}{MT} \sum_{i=1}^{M} \sum_{t=1}^{T} \mathbb{1}(\text{consumer } i \text{ visits the shop in period } t) \\
&= \frac{1}{MT} \sum_{i=1}^{M} \sum_{t=1}^{T} \max \left\{ \sum_{j=0}^{J_1} Y_{ijt}^1, ..., \sum_{j=0}^{J_K} Y_{ijt}^K \right\}.
\end{aligned}$$

Recall that $\mathbf{h}_i^k = \{(Y_{i01}^k, ..., Y_{iJ1}^k), ..., (Y_{i0T}^k, ..., Y_{iJT}^k)\}$ represents the purchase history of consumer i for category k as previously, and $Y_{ijt}^k = 1$ if consumer i chose product j at time t and zero otherwise. Also, note that $\sum_{j=0}^{J_k} Y_{ijt}^k = 0$ if consumer i did not have a purchase occasion in period t (i.e., he chose none of the products or the outside option). We do not observe consumers that did not get a single shopping occasion at our representative store, hence our sample of consumers is a selected sample. In practice, this means that $\mathbb{E}(\tilde{\rho}) > \rho$. We can correct for the bias by accounting for the missing consumers. We compute the following unbiased estimator

$$\begin{aligned}
\widehat{\rho} &= \frac{1}{\widetilde{M}T} \sum_{i=1}^{M} \sum_{t=1}^{T} \max \left\{ \sum_{j=0}^{J_1} Y_{ijt}^1, ..., \sum_{j=0}^{J_K} Y_{ijt}^K \right\} \\
&= \left(\frac{M}{\widetilde{M}} \right) \cdot \widetilde{\rho}
\end{aligned}$$

where $\widetilde{M}$ is the market size including consumers who never shop at the store.

CASE 2: Heterogenous shopping probabilities. We now consider the case where consumers have heterogenous shopping probabilities ρ_i that do not depend on their type α_i. In this case the market shares are

$$
\begin{aligned}
s_j &= \int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\Omega} \mathbb{1}_S(\omega_i) s_j(\alpha_i) dF_{\omega,\alpha}(\omega_i, \alpha_i) \\
&= \int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\Omega} \mathbb{1}_S(\omega_i) s_j(\alpha_i) dF_{\omega}(\omega_i) dF_{\alpha}(\alpha_i) \\
&= \left(\int_{\Omega} \mathbb{1}_S(\omega_i) dF_{\omega}(\omega_i) \right) \cdot \left(\int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha_i) dF_{\alpha}(\alpha_i) \right) \\
&= \underbrace{\frac{1}{M} \sum_{i=1}^{M} \rho_i}_{\widetilde{\rho}} \cdot \left(\int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha_i) dF_{\alpha}(\alpha_i) \right).
\end{aligned}
$$

In this case, we see that the model with consumer heterogeneity collapse to the simple model with homogeneous shopping probabilities. This follows from independence between ω_i and α_i.

CASE 3: Dependence on price sensitivity. We now discuss the case where consumers shopping probabilities are not independent of their type. We consider without loss of generality the case where the random variable $\rho_i = \rho(\alpha_i)$, but a similar argument can be made by reversing the relationship. We can write the market share as

$$
\begin{aligned}
s_j &= \int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\Omega} \mathbb{1}_S(\omega_i) s_j(\alpha_i) dF_{\omega,\alpha}(\omega_i, \alpha_i) \\
&= \int_{\underline{\alpha}}^{\overline{\alpha}} \int_{\Omega} \mathbb{1}_S(\omega_i) s_j(\alpha_i) dF_{\omega|\alpha}(\omega_i|\alpha_i) dF_{\alpha}(\alpha_i) \\
&= \int_{\underline{\alpha}}^{\overline{\alpha}} \left(\int_{\Omega} \mathbb{1}_S(\omega_i) dF_{\omega|\alpha}(\omega_i|\alpha_i) \right) s_j(\alpha_i) dF_{\alpha}(\alpha_i) \\
&= \int_{\underline{\alpha}}^{\overline{\alpha}} \rho(\alpha) s_j(\alpha_i) dF_{\alpha}(\alpha_i) \\
&\neq \rho \int_{\underline{\alpha}}^{\overline{\alpha}} s_j(\alpha_i) dF_{\alpha}(\alpha_i)
\end{aligned}
$$

since $\rho(\alpha_)$ is not independent of α_i. Thus, for our estimation to be valid, we can allow for heterogeneous shopping probabilities, but not dependence with consumers types.