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

Isis Durrmeyer† Jean-François Fournel‡ Mario Samano§

October 6, 2025

Abstract

Our analysis uses data on prices, aggregate quantities, and individual purchase histories from a large supermarket chain in the U.S. and an empirical model to represent grocery shopping by consumers and the supermarket pricing strategies. We estimate demand for 24 product categories and recover 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 that 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 histories and the consequences for consumer surplus. Our results show that price personalization leads to an increase in profits of around 4% in all categories. We find that the effect on consumers is mostly redistributive, with a small number of consumers experiencing large losses and a large number of consumers experiencing small gains.

Keywords: personalized pricing, differentiated goods, 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 Hester Zhang and 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 can set prices based on the sequence of past purchase histories as shown by 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 through price personalization. The fundamental economic question behind the profitability of this type 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. We then apply this model to grocery purchases to quantify the welfare consequences of behavior-

[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.

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 24 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, 2001a, 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 the demand for a large set of product categories is Döpper et al. (2023).

Once we have estimated 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.[4] 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's 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

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

Shiller, 2020, for instance). In comparison, we explicitly model the supply side and capture 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, by around 4% over all categories. The gains monotonically increase with the purchase history's length and are concave (in most categories) with history length, indicating decreasing marginal returns to information from purchase histories. The gains in profits at the category level are highly asymmetric. We find that for some product categories such as bottled water, price personalization can lead to as much as 9% more profits than setting uniform prices. For other categories such as laundry detergent or dry pet food, setting personalized prices has a negligible effect.

On the consumer side, we find that a large number of consumers are presented with slightly infra-uniform prices and a small group of price-insensitive consumers is presented with supra-uniform prices. The overall effect on consumer surplus is negative: the small gains by the many price-sensitive consumers do not completely offset the large losses by the few price-insensitive consumers.

We define the value of information, or the value of the purchase histories, as the supermarket's willingness-to-pay to acquire consumers individual purchase histories, and the compensation that is required to make each consumer as well off under price personalization and uniform pricing. We find that the supermarket has a positive willingness-to-pay for all consumer histories: either it extracts more profits by increasing price or by expanding demand (when it offers discounts). Around 18% of consumers require a compensation for them to accept to share their data, e.g., by opting in a loyalty program that allows the seller to track their purchases over time. In some cases, their data is profitable enough that the seller could compensate these consumers and generate societal gains. In other cases, the compensation required can be as much as three times the willingness-to-pay of the seller. For the remaining 82% of consumers, they gain from sharing their personal data, and thus price personalization leads to win-win interactions.

As a final exercise, we propose a two-stage pricing game, where consumers are free to opt in or out of a loyalty program that enables personalized pricing. The supermarket proposes a transfer to consumers to acquire their consumer data, and then personalize the prices for the consumers that opt in. We find that the optimal transfer for the supermarket is to charge a small membership fee to adhere to the loyalty program. Consumers that adhere receive a personalized discount over the uniform price, and consumer adhere based on the size of the membership fee, and the potential savings to be realized at the supermarket.

Relation to the literature. Our demand and supply models are based on the classical setup of Berry (1994) and Berry et al. (1995), but with crucial differences regarding how new information is used to determine optimal discriminatory prices. Most of the literature listed below also uses the same discrete choice setup.

This paper is directly related to the implementation of personalized pricing for a multi-product monopolist. Although in our case we use Bayesian updating to uncover consumer types without a hierarchical model, there is a well-established literature on hierarchical Bayesian models for demand estimation that can be traced to Rossi et al. (1996). Among other differences, our model covers a much larger set of goods and contains a supply side. More recent applications of Bayesian updating under monopoly include Shin et al. (2012) and Dubé and Misra (2023), which, in addition, rely on experimental data. Other approaches in the monopolistic case include Waldfogel (2015).[5]

In our empirical application, we assume that the monopolist has implemented uniform pricing in the past. This pricing strategy in supermarket chains has been the object of several recent studies across different chains and across different markets (DellaVigna and Gentzkow (2019), Hitsch et al. (2019), Chandra and Lederman (2018), Puller and Taylor (2012), and Eizenberg et al. (2021)).

Our approach to estimating demand systems leverages the repeated interactions between the supermarket and participants in a loyalty program to categorize consumers into groups, based on the sequence of their purchases. We estimate fundamental parameters that characterize their preferences and compute the share of consumers that belong to each group for each product category. Our demand system estimation is most closely related to the estimator in Fox et al. (2016).[6] We improve on their methodology by including market share constraints

[5]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.

[6]See Remark 1 of Example 1.

to the optimization problem, similarly to the MPEC specification in Dubé et al., 2012, which helps with the identification of preference parameters and improves the overall performance of the estimator.

Some recent examples of demand estimation using supermarket product data include Döpper et al. (2023) , Compiani (2022), Thomassen et al. (2017) , and Smith et al. (2022). As explained above, this paper differentiates from this literature in that we can assess the goodness-of-fit of our model by a simple comparison of the recovered marginal costs and the observed marginal costs, it combines Bayesian updating rules with a frequentist estimator, and it covers a large set of products across different categories.

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

2 The Model

We describe a model of demand and supply in which a monopolistic, multi-product seller has repeated interactions with a set of heterogeneous consumers. Initially, the seller knows the shape of the distribution of consumers' preferences in the population, but not individual-specific preferences. The seller observes consumers' purchases over time which reveal some information about each consumer's willingness to pay for each product. The seller then uses this information to personalize prices for these consumers on their next visit to the store.[7]

We contrast this pricing strategy (based on past purchase histories) with two benchmarks: first, the case where the seller does not learn about consumers beyond the distribution of preferences in the population and is restricted to set uniform prices, second, the case where it has full information on consumers and performs first-degree price discrimination.

[7]Throughout, we assume that agents play a repeated static game. Consumers are myopic: they do not consider future interactions with the seller when making purchase decisions. Similarly, the seller does not experiment with prices in an effort to gather more information from consumers. We leave these dynamic considerations for future work.

Demand. We consider that demand arises from a heterogeneous set of consumers, denoted by $i = 1, \ldots, M$. Each period $t = 1, \ldots, T$, consumers have an exogenous, consumer-specific probability to visit the store to incur a purchase occasion.[8] We denote this probability as

$$\rho_i = \Pr(\text{Purchase occasion}).$$

Conditionally on having a purchase occasion, consumers choose to purchase one of the $j = 1, \ldots, J$ products available or to purchase nothing (the outside option) which we denote by $j = 0$. Whenever consumers do not have a purchase occasion, they are restricted to choosing the outside option. In this case, the product they would have chosen given a purchase occasion is unobserved by the seller.

We assume that consumers have heterogeneous preferences for the available products. The indirect utility of consumer i buying product j in a given purchase occasion is

$$u_{ij} = \delta_{ij}(\beta_i) + \alpha_i p_{ij} + \epsilon_{ij},$$

where to keep the number of indices reasonable, we do not write the purchase occasion index. We denote by p_{ij} the price set for consumer i for product j (which may vary with i as a result of price discrimination), α_i is the price sensitivity parameter, and δ_{ij} is a quality index which depends on β_i, a vector of consumer-specific preference parameters.[9] The utility of the outside option is normalized such that

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

Throughout, we refer to the collection of consumers' preferences parameters and their purchase occasion probability as the consumer's type, denoted $\Theta_i = (\alpha_i, \beta_i, \rho_i)$, with joint cumulative distribution function $F_\Theta(\Theta_i)$.

[8] In practice, we could relax the exogeneity assumption by, for example, allowing the probability to obtain a purchase occasion to depend on the prices offered by the seller. In oligopolistic settings with multiple stores, a valid strategy for sellers is to offer large discounts on specific products to attract consumers, provided that it is costly for consumers to visit several stores (see for example Thomassen et al., 2017 on competition among supermarkets). In this case, ρ_i would be endogenous and depend on the price level at each store. To keep the model tractable, we focus on the case with a single seller, and we assume that consumers are not aware of the price level offered by the seller when deciding whether or not to visit the store.

[9] Several different parametrizations for δ_{ij} are possible. One common example is $\delta_{ij} = X_j\beta + \xi_j + \sum_r X_{jr}\sigma_r\nu_{ir}$, where the $X_j = (X_{j1}, \ldots, X_{jR})$ are exogenous product characteristics, ξ_j is an unobserved quality index, $\nu_i = (\nu_{i1}, ..., \nu_{iR})$ includes consumer demographic characteristics and random coefficients, and $\beta_i = (\beta_{i1}, ..., \beta_{iR})$, $\beta_{ir} = \beta_r + \sigma_r\nu_{ir}$ are consumer specific preference parameters (see Nevo, 2001b). In the application, we use the following parametrization, $\delta_{ij}(\beta_i) = \delta_j + \beta_i$, i.e., the sum of a product fixed effect and a consumer-specific fixed effect.

We make the standard assumption that the ϵ_{ij} are independent and identically distributed across individuals and products and follow an extreme value distribution. Under this distributional assumption, the probability that a consumer with type Θ_i purchases product j during a purchase occasion is

$$\begin{aligned}\mathbb{P}_j(\theta_i) &= \Pr(u_{ij} > u_{ij'}, \forall j' \neq j \mid \text{Purchase occasion}) \\ &= \frac{\exp\left(\delta_{ij}(\beta_i) + \alpha_i p_{ij}\right)}{1 + \sum_{j'} \exp\left(\delta_{ij'}(\beta_i) + \alpha_i p_{ij'}\right)},\end{aligned}$$

where $\theta_i = (\alpha_i, \beta_i)$ are the consumer's preference parameters (excluding ρ_i). The joint probability that a consumer with type Θ_i gets a purchase occasion and chooses product j is

$$\begin{aligned}s_j(\Theta_i) &= \Pr(u_{ij} > u_{ij'}, \forall j' \neq j \mid \text{Purchase occasion}) \times \Pr(\text{Purchase occasion}) \\ &= \mathbb{P}_j(\theta_i)\, \rho_i\end{aligned}$$

Finally, the overall market share of product j is

$$\begin{aligned}s_j &= \int s_j(\Theta_i)\, dF_\Theta(\Theta_i), \\ &= \int \mathbb{P}_j(\theta_i)\, \rho_i\, dF_\Theta(\Theta_i).\end{aligned} \tag{1}$$

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

$$CS(\Theta_i) = -\frac{\rho_i}{\alpha_i} \ln\left(1 + \sum_{j=1}^{J} \exp(\delta_{ij}(\beta_i) + \alpha_i p_{ij})\right),$$

where p_{ij} is the price faced by consumer i for product j (which could depend on consumers characteristics or their past purchases), α_i is, as previously noted, the price sensitivity parameter, β_i is a vector of consumer preference parameters, and ρ_i is the probability to obtain a purchase occasion.[10] The aggregate consumer surplus is therefore

$$
\begin{aligned}
CS &= \int CS(\Theta_i)\, dF_\Theta(\Theta_i), \\
&= -\int \frac{\rho_i}{\alpha_i} \ln\left(1 + \sum_{j=1}^{J} \exp(\delta_{ij}(\beta_i) + \alpha_i p_{ij})\right) dF_\Theta(\Theta_i).
\end{aligned}
$$

Uniform pricing. Throughout, we assume that the seller sets prices, not the underlying manufacturers of products. Therefore, we consider pricing for a multi-product monopolist. To simplify the notation, we let J represent both the total number of products and the set of all products. Under uniform pricing, the seller sets one price per product for all consumers. This could arise if the seller does not have access to the technology necessary to implement price personalization or does not have information about individual consumer preferences.

Let π^U denote the expected profits under uniform pricing,

$$
\pi^U = \sum_{j \in J} (p_j - c_j) \cdot \int s_j(\Theta_i)\, dF_\Theta(\Theta_i) \cdot M,
$$

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 seller sets prices such that the following first-order conditions are satisfied,

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

where $\mathbf{s}$, $\mathbf{p}$, and $\mathbf{c}$ are the vectors of market shares, prices, and marginal costs, respectively. Each element of $\mathbf{s}$ is equal to $\int \mathbb{P}_j(\theta_i)\, \rho_i\, dF_\Theta(\Theta_i)$. The matrix $\mathbf{\Lambda}$ contains all the derivatives of the market shares such that the element (j, j') of $\mathbf{\Lambda}$ is equal to

$$
\Lambda_{jj'} = \begin{cases} \int \alpha_i\, \mathbb{P}_{j'}(\theta_i)\, (1 - \mathbb{P}_{j'}(\theta_i))\, \rho_i\, dF_\Theta(\Theta_i) & \text{if } j = j', \\ -\int \alpha_i\, \mathbb{P}_j(\theta_i)\, \mathbb{P}_{j'}(\theta_i)\, \rho_i\, dF_\Theta(\Theta_i) & \text{if } j \neq j'. \end{cases}
$$

Note that $\mathbf{\Lambda}$ is a function of $\mathbf{p}$ through the market share expressions $\mathbb{P}_j(\theta_i)$. The solution to this system of non-linear equations is the vector of uniform prices $\mathbf{p}^U$.

[10] Consumer surplus at the consumer level can be written as

$$
CS(\Theta_i) = \rho_i \cdot CS(\theta_i \mid \text{Purchase occasion}) + (1 - \rho_i) \cdot CS(\theta_i \mid \text{No purchase occasion}),
$$

where $CS(\theta_i \mid \text{Purchase occasion})$ follows the usual log-sum formula and $CS(\theta_i \mid \text{No purchase occasion})$ is normalized to zero.

Firm's information acquisition process. The seller observes the realizations of the purchase occasions and, when a purchase occasion occurs, it observes the choices made by each individual. The price environment is always known to the seller, irrespectively of purchase occasions. We call the sequence of individual choices for T periods consumer histories. We assume that the seller forms an expectation about consumers' types and learn from observing purchase histories using Bayes's rule.

Denote by $\mathbf{h}_\ell = \{(Y_{\ell 01}, \ldots, Y_{\ell J1}), \ldots, (Y_{\ell 0T}, \ldots, Y_{\ell JT})\}$ the purchase history ℓ of length T, where $Y_{\ell jt}$ is equal to 1 if the consumer bought 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. This occurs whenever a consumer does not have a purchase occasion.

There is a key difference between choosing the outside option during a purchase occasion and not having a purchase occasion. In the first case, the seller observes a choice, given prices, and is able to update its belief about the consumer's preference parameters $\theta_i = (\alpha_i, \beta_i)$. In the second case, the data are censored and the seller does not observe which product the consumer would have chosen given prices.

Interestingly, some information about the consumer's willingness to pay for products can still be transferred to the seller if, for example, the preference parameters θ_i are correlated with the purchase occasion probability ρ_i. In this case, not visiting the store provides some information to the seller on ρ_i, which in turn is informative of θ_i, which matters for price personalization. If, on the other hand, θ_i and ρ_i are independent, then learning about ρ_i does not provide any information on θ_i. In this case, the seller only updates its belief about θ_i during shopping occasions.

We index histories by $\ell = 1, \ldots, L$ to differentiate the history's index from the consumer's index, i, as multiple consumers may share the same purchase history. Finally, we denote by $M_{\mathbf{h}_\ell} \geq 0$ the number of potential consumers with purchase history $\mathbf{h}_\ell$.[11] We have $M_{\mathbf{h}_1} + \ldots + M_{\mathbf{h}_L} = M$. Following Bayes's rule, we can write

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

[11] 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 are never observed by the seller and $M_{\mathbf{h}_\ell} = 0$. From a practical point of view, the seller does not have to solve for all price vectors that arise from all possible histories, it only need to provide prices for the observed histories. Solving for all prices would impose a heavy computational burden on the seller (and on us while solving for counterfactual experiments).

The probability of observing history $\mathbf{h}_\ell$ conditional of Θ_i is

$$f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i) = \prod_{t=1}^{T}\left[\prod_{j=0}^{J}\mathbb{P}_{jt}(\theta_i)^{Y_{\ell jt}}\right]\rho_i^{\sum_{j=0}^{J}Y_{\ell jt}}(1-\rho_i)^{1-\sum_{j=0}^{J}Y_{\ell jt}}, \tag{4}$$

because demand shocks are independent across purchase occasions. The probability of observing purchase history $\mathbf{h}_\ell$, unconditionally of the consumer's type Θ_i can be written as

$$\begin{aligned} f_{\mathbf{h}_\ell}(\mathbf{h}_\ell) &= \int f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i)\,dF_\Theta(\Theta_i), \\ &= \int \prod_{t=1}^{T}\left[\prod_{j=0}^{J}\mathbb{P}_{jt}(\theta_i)^{Y_{\ell jt}}\right]\rho_i^{\sum_{j=0}^{J}Y_{\ell jt}}(1-\rho_i)^{1-\sum_{j=0}^{J}Y_{\ell jt}}\,dF_\Theta(\Theta_i). \end{aligned} \tag{5}$$

Finally, following Bayes's rule, equation (3) for the distribution of consumer types conditional on observing purchase history $\mathbf{h}_l$ is

$$f_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell) = \frac{\prod_{t=1}^{T}\left[\prod_{j=0}^{J}\mathbb{P}_{jt}(\theta_i)^{Y_{\ell jt}}\right]\rho_i^{\sum_{j=0}^{J}Y_{\ell jt}}(1-\rho_i)^{1-\sum_{j=0}^{J}Y_{\ell jt}}f_\Theta(\Theta_i)}{\int\prod_{t=1}^{T}\left[\prod_{j=0}^{J}\mathbb{P}_{jt}(\theta_i)^{Y_{\ell jt}}\right]\rho_i^{\sum_{j=0}^{J}Y_{\ell jt}}(1-\rho_i)^{1-\sum_{j=0}^{J}Y_{\ell jt}}dF_\Theta(\Theta_i)}. \tag{6}$$

Personalized pricing. Under personalized pricing, the seller sets prices $\mathbf{p}^H = \mathbf{p}(\mathbf{h}_\ell)$ to consumers with purchase history $\mathbf{h}_\ell$. The profits function can be expressed as

$$\pi^H = \sum_{\ell=1}^{L}\sum_{j\in J}(p_j(\mathbf{h}_\ell)-c_j)\cdot\int s_j(\Theta_i)\,dF_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell)\cdot M_{\mathbf{h}_\ell},$$

where $M_{\mathbf{h}_\ell}$ denotes the number of consumers with purchase history $\mathbf{h}_\ell$. The seller sets a different menu of prices for each purchase history $\mathbf{h}_\ell$ such that the following system of J nonlinear first-order conditions is satisfied,

$$\mathbf{s}(\mathbf{h}_\ell)+\mathbf{\Lambda}(\mathbf{h}_\ell)\,(\mathbf{p}(\mathbf{h}_\ell)-\mathbf{c}) = 0, \quad \forall \ell = 1,...,L, \tag{7}$$

giving a total of $J \times L$ equations. Each element of $\mathbf{s}(\mathbf{h}_\ell)$ is the vector of market shares, computed using the conditional density obtained through Bayes's rule above,

$$s_j(\mathbf{h}_\ell) = \int \mathbb{P}_j(\theta_i)\,\rho_i\,dF_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell).$$

Similarly, each element (j,j') of the matrix $\mathbf{\Lambda}(\mathbf{h}_\ell)$ can be expressed as

$$\Lambda_{jj'}(\mathbf{h}_\ell) = \begin{cases} \int \alpha_i\,\mathbb{P}_{j'}(\theta_i)\,(1-\mathbb{P}_{j'}(\theta_i))\,\rho_i\,dF_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell) & \text{if } j = j', \\ -\int \alpha_i\,\mathbb{P}_j(\theta_i)\,\mathbb{P}_{j'}(\theta_i)\,\rho_i\,dF_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell) & \text{if } j \neq j'. \end{cases}$$

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

$$\pi^P = \sum_{i=1}^{M} \sum_{j \in J} (p_j(\theta_i) - c_j)\, s_j(\Theta_i).$$

The seller sets prices $\mathbf{p}^P = \mathbf{p}(\theta_i)$ for each consumer type separately. The optimal price for a type Θ_i consumer is such that the following J first-order conditions are satisfied,

$$\mathbb{P}(\theta_i) + \mathbf{\Lambda}(\theta_i)(\mathbf{p}(\theta_i) - \mathbf{c}) = 0, \quad \forall i = 1, ..., M, \tag{8}$$

where each element of $\mathbb{P}(\theta_i)$ is equal to $\mathbb{P}_j(\theta_i)$, and $\mathbf{\Lambda}(\theta_i)$ contains all the derivatives of $\mathbb{P}(\theta_i)$ such that the element (j, j') of $\mathbf{\Lambda}(\theta_i)$ is equal to

$$\Lambda_{jj'}(\theta_i) = \begin{cases} \alpha_i\, \mathbb{P}_{j'}(\theta_i)\,(1 - \mathbb{P}_{j'}(\theta_i)) & \text{if } j = j', \\ -\alpha_i\, \mathbb{P}_j(\theta_i)\, \mathbb{P}_{j'}(\theta_i) & \text{if } j \neq j'. \end{cases}$$

Note that $\mathbb{P}(\theta_i)$ and $\mathbf{\Lambda}(\theta_i)$ do not depend on ρ_i, so prices under perfect price discrimination, $\mathbf{p}^P = \mathbf{p}(\theta_i)$, are independent from the probability to obtain a shopping occasion ρ_i.

3 Estimation and Identification

We assume that the observed market outcome is the consequence of a monopolistic, multi-product seller setting uniform prices, and taking wholesale prices and retail costs as given. We observe consumers' purchases over $t = 1, \ldots, T$ periods. In our application, we parametrize the quality index as

$$\delta_{ijt}(\beta_i) = \delta_{jt} + \beta_i,$$

that is, the sum of a product-by-period fixed effect, δ_{jt}, and a consumer-specific preference parameter, β_i. We assume throughout that the product fixed effects (or mean utilities) are independent of the consumers' preference parameters.

In addition, we assume that the distribution of consumer types is discrete, with D known points of support.[12] Using a finite number of discrete consumer types is very appealing in our

[12] Fox et al. (2016) establish the consistency and the rates of convergence of these types of fixed grid, nonparametric estimators.

context as it allows both for rich consumer heterogeneity and keeping the price discrimination dimension tractable in counterfactual analysis.

The known points of support of the distribution of consumers' preferences, $\{\Theta_d\}_{d=1,\ldots,D} = \{\alpha_d, \beta_d, \rho_d\}_{d=1,\ldots,D}$, are taken as given and the estimated parameters are the probability mass function over the fixed consumer types, $\{\phi_d\}_{d=1,\ldots,D}$, and the product fixed effects $\{\boldsymbol{\delta}_t\}_{t=1,\ldots,T}$, with $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{Jt})$. We estimate these parameters by maximum likelihood, under the constraints that the model's predicted market shares, $\sum_d \phi_d \, s_{jt}(\Theta_d, \boldsymbol{\delta}_t)$, equal the observed market shares, $\mathfrak{s}_{jt}$.

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 $\mathbf{h}_i = \{Y_{ijt}\}_{j=0,\ldots,J;\, t=1,\ldots,T}$, where, as noted previously, Y_{ijt} is an indicator variable that identifies which product was chosen by consumer i in period t.[13] The contribution of consumer i to the likelihood function is

$$\Pr(\mathbf{h}_i) = \sum_{d=1}^{D} \phi_d \prod_{t=1}^{T} \left[\prod_{j=0}^{J} \mathbb{P}_{jt}(\theta_d, \boldsymbol{\delta}_t)^{Y_{ijt}} \right] \rho_d^{\sum_{j=0}^{J} Y_{ijt}} \left(1 - \rho_d\right)^{1 - \sum_{j=0}^{J} Y_{ijt}},$$

which we derived in the previous section, and

$$\mathbb{P}_{jt}(\theta_d, \boldsymbol{\delta}_t) = \frac{\exp(\delta_{jt} + \beta_d + \alpha_d p_{jt})}{1 + \sum_{j'} \exp(\delta_{j't} + \beta_d + \alpha_d p_{j't})}$$

is the probability that consumer i chooses product j conditional on having a purchase occasion in period t. The objective function consists of the log-likelihood of the sample,

$$\mathscr{L}(\{\phi_d\}_{d=1,\ldots,D}, \{\boldsymbol{\delta}_t\}_{t=1,\ldots,T}) = \sum_{i=1}^{M} \ln\left(\Pr(\mathbf{h}_i)\right), \tag{9}$$

where the sum aggregates over consumers, possibly with some consumers sharing the same purchase history. We include a set of constraints based on the market share equation (1), that is,

$$\mathfrak{s}_{jt} = \sum_{d=1}^{D} \phi_d \, \frac{\rho_d \cdot \exp(\delta_{jt} + \beta_d + \alpha_d p_{jt})}{1 + \sum_{j'} \exp(\delta_{j't} + \beta_d + \alpha_d p_{j't})}, \; \forall j, \, \forall t, \tag{10}$$

[13]Recall that, whenever consumers receive a purchase occasion, they must choose one of the $j = 1, \ldots, J$ products available or the outside option denoted by $j = 0$. When they do not receive a purchase occasion, we do not observe which choice the consumer would have made given the pricing environment and we have $Y_{ijt} = 0$ for all $j = 0, \ldots, J$.

where $\mathscr{s}_{jt}$ is the observed market share. For each period and for given values of $\{\phi_d\}_{d=1,\ldots,D}$ and $\{\alpha_d, \beta_d, \rho_d\}_{d=1,\ldots,D}$, this system of equations defines a unique vector of mean utilities $\{\boldsymbol{\delta}_t\}_{t=1,\ldots,T}$ as shown by Berry (1994).

The estimation is based on maximizing the likelihood function (9), subject to the nonlinear constraints given by equation (10) and the estimated probabilities $\{\phi_d\}_{d=1,\ldots,D}$ constituting a valid probability mass function. Putting everything together, the problem to solve is

$$
\begin{aligned}
&\max_{\{\phi_d\}_{d=1,\ldots,D},\,\{\boldsymbol{\delta}_t\}_{t=1,\ldots,T}} \mathscr{L}(\{\phi_d\}_{d=1,\ldots,D},\, \{\boldsymbol{\delta}_t\}_{t=1,\ldots,T}),\\
&\text{s.t.} \quad \mathscr{s}_{jt} = \sum_{d=1}^{D} \phi_d \, \frac{\rho_d \cdot \exp(\delta_{jt} + \beta_d + \alpha_d p_{jt})}{1 + \sum_{j'} \exp(\delta_{j't} + \beta_d + \alpha_d p_{j't})}, \qquad (11)\\
&\qquad \sum_{d=1}^{D} \phi_d = 1, \quad \phi_d \geq 0.
\end{aligned}
$$

Our estimation method combines elements from the nonparametric estimator described in Fox et al. (2016) (see Example 1, Remark 1), which relies on a fixed grid characterization of the demand estimation optimization problem, and the MPEC formulation of the BLP model (see, Dubé et al., 2012), which relies on a constrained optimization to recover the unobserved quality index (here, the δ_{jt}).

4 Data and Estimation Results

4.1 Data

We estimate our model using data on grocery purchases provided by DecaData.[14] These data contain daily point-of-sale transactions from 853 stores in 13 U.S. states, retailer-level product deliveries and restocking, and detailed product information, including quantities transacted, retail prices, and wholesale prices. All sales are recorded daily, for each transaction, at the universal product code level (henceforth UPC). A key feature of the data is that it contains a consumer identifier for a large subset of consumers enrolled in a loyalty program. This allows us to track their purchases over time, reconstructing their purchase histories from January 1 to December 31, 2018.[15]

[14] Source: https://decadata.io/data.

[15] The consumer identifiers are store specific and we cannot track consumers purchases across stores.

Our empirical application relies on consumers' past purchases to reveal their willingness to pay for products which allows price personalization. Due to data limitations, we do not use demographics to reveal consumer types or to compute personalized prices. 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 focus on one representative supermarket. We choose a store near the median store in terms of total revenues. Our chosen store had \$13.1 million in sales in 2018, placing it in the 52nd percentile of the distribution. At the same time, we restrict our analysis to the 17,756 consumers with a loyalty card who made at least one purchase at our representative store in 2018. Sales from these consumers reached \$10.4 million in 2018, approximately 80% of all sales recorded at our representative store. The sample of consumers we can track over time was not selected at random so we cannot completely rule out selection issues. However, we cannot track the excluded consumers over time and recover their purchase histories, which is crucial for studying personalized pricing in our context. Our results should be carefully interpreted with this caveat in mind.

We restrict our attention to 24 product categories within 6 food departments. The chosen departments are alcohol, dairy, frozen food, groceries, packaged meat, and taxable (non-food) groceries. The list of chosen categories is available in Table 1. We consider each category $k = 1, \ldots, 24$ to be a separate market and we ignore the potential complementarities between products 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.

We assume that each manufacturer active in a given category produces a separate, differentiated product in that category. We convert all UPC in a category in standardized units, based on the most commonly sold package size, and we aggregate over UPC sold by each manufacturer to form the set of products available in each market. Sales are aggregated at the consumer-month level, and we define each month to be a potential shopping occasion. Approximately 38% of consumers visit the store in any given months and consumers visit the store on average 1.5 times per month (the median number of visits per month is 0).

Table 1 summarizes the product characteristics of the final sample. Additional details on the data are relayed to Appendix B. 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. All prices are set in each category by

Table 1: Summary statistics of the final sample.

Category	# Products	Standard package size	Avg. monthly quantity (per consumer)	Avg. unit price ($)	Avg. unit cost ($)	Avg. monthly sales ($)
Alcohol						
Beer	4	144 oz	1.85	13.19	10.27	5,862.92
Wines	7	25.4 oz	3.73	7.39	5.47	3,781.15
Dairy						
Butter/margarine/spreads	5	16 oz	2.1	2.29	1.60	4,981.65
Cheese	4	8 oz	3.9	2.55	1.58	15,464.01
Creams/creamers	4	32 oz	1.69	4.15	2.67	2,676.66
Fresh eggs	2	12 ct	2.28	2.82	1.94	5,478.81
Refrigerated juice/beverage	4	59 oz	3.7	2.41	1.71	7,517.17
Milk	5	64 oz	4.13	2.65	1.74	14,843.78
Yogurt	4	5.3 oz	9	0.79	0.55	2,671.83
Frozen food						
Ice cream	4	48 oz	2.24	4.55	3.03	4,220.81
Frozen pizza	4	10 oz	5.44	1.73	1.29	2,651.10
Frozen potatoes/onions	4	32 oz	2.01	2.87	2.07	3,267.59
Frozen vegetables	4	12 oz	5.34	1.56	0.98	7,118.59
Grocery						
Cereal	2	12.5 oz	6.12	18.36	16.02	4,858.60
Coffee	5	12 oz	2.42	3.62	2.94	4,332.17
Crackers	6	12 oz	1.26	10	7.83	2,996.04
Snacks grocery	5	12.4 oz	2.04	3.03	2.40	3,682.44
Soft drinks/mixers	3	8 oz	3.93	2.86	2.17	18,334.71
Water	5	144 oz	4.34	7.65	4.82	24,503.03
Packaged meat						
Bacon	6	128 oz	5.23	4.13	2.29	8,894.04
Dinner sausage	6	24 oz	1.17	9.54	6.78	7,072.76
Taxable grocery						
Bleach/stain removers	6	14 oz	4.11	2.81	1.81	11,324.01
Laundry detergent	4	128 oz	2.59	3.38	2.54	1,417.80
Pet food	4	32 oz	2.33	6.61	5.52	1,538.97

Notes: "Standard package size" is the most common package size sold in a given category. "Average monthly quantity" is the average quantity purchased per consumer, in standard units, conditionally on purchasing any product in the category in a given month. "Average unit price" is the weighted average retail price for one standard unit and "Average unit cost" is the weighted average wholesale price for one standard unit. "Average monthly sales" is the total sales by category and month, in dollars, averaged over months.

the supermarket, treating each category as a separate multi-product pricing problem.

4.2 Purchase histories and market shares

We next describe how we construct consumers' 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. Let $k = 1, \ldots, K$ index categories and $j = 1, \ldots, J_k$ index the products in category k. We expand the notation from the previous chapter such that consumer i's purchase history of length T is denoted as

$$\mathbf{h}_i^T = \{(Y_{i01}^k, ..., Y_{iJ1}^k), ..., (Y_{i0T}^k, ..., Y_{iJT}^k)\}_{k=1,...,K},$$

where Y_{ijt}^k is equal to 1 if product j from category k is chosen in purchase occasion t and 0 otherwise. Let q_{ijt}^k be the quantity of product j from category k purchased by consumer i in shopping occasion t, expressed in normalized units as per Table 1. We compute the entries in each purchase history as follows,[16]

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

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

$$Y_{i0t}^k = \begin{cases} 1 & \text{if } q_{ijt}^k = 0, \forall j > 0 \text{ and ``Purchase occasion''} \\ 0 & \text{if } q_{ijt}^k = 0, \forall j > 0 \text{ and ``No purchase occasion''} \end{cases}.$$

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

$$\mathfrak{s}_{jt}^k = \frac{1}{M} \sum_{i=1}^{M} Y_{ijt}^k, \;\; \forall j = 0, 1, ..., J_k.$$

[16] In case of a tie, we assign one of the products involved in the tie randomly. Ties did not occurs in the data.

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. Our final set of consumers includes 17,756 individuals. 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 additional normalization.

4.3 Estimation results

We perform a separate estimation on each product category, assuming uniform prices, and taking consumers' purchase histories as given. Preference parameters are consumer-category specific, meaning that a typical consumer is characterized by a vector of parameters $\Theta_i = (\alpha_i^1, ..., \alpha_i^K, \beta_i^1, ..., \beta_i^K, \rho_i)$. Notice that ρ_i is consumer specific and does not vary by category.

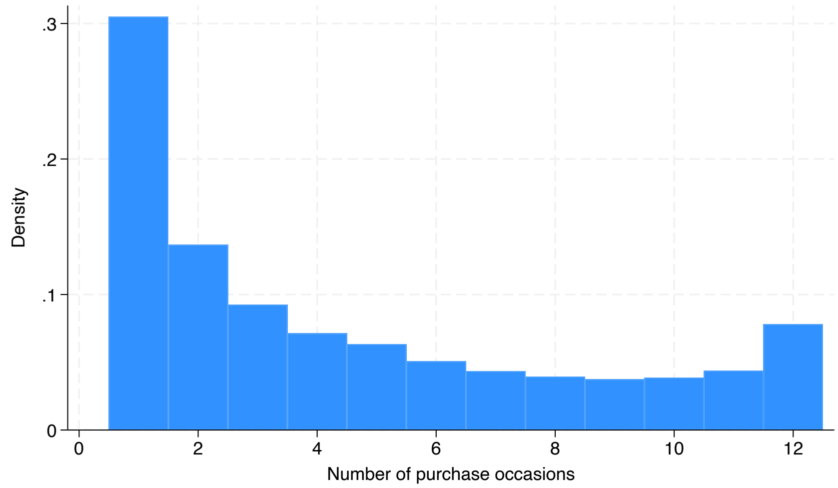
The estimation is performed in three stages. First, we recover the support of ρ_i by fitting a finite mixture model on the distribution of realized purchase occasions. Second, we estimate a flexible version of the model presented in Section 3, equation (11), to recover the support of consumers' preferences parameters, α_i^k and β_i^k, in each product category. Finally, we take the estimated supports of $\Theta_i^k = (\alpha_i^k, \beta_i^k, \rho_i)$ as given and estimate (11) on a fixed grid that lies inside the support.

We provide more details about each step of the estimation in what follows. Additional details are available in Appendix C.

Purchase occasions. For each consumer in our sample, we observe purchases made at the representative store in each month (across all categories, including the excluded categories). We assume that consumers who purchased nothing in a given month did not obtain a purchase occasion in that month. Conversely, consumers with positive purchases in a given month visited the store that month. Figure 1 presents the distribution of purchase occasions, cumulative over 12 months according to this definition. The fact that the graph is U-shaped suggests that there is significant heterogeneity in the probabilities to obtain a purchase occasion.[17]

[17] The relevant distribution for the number of purchase occasion over 12 months is the Binomial($N = 12, p = \rho_i$). For $\rho_i = 0.377$ and 17,756 observations, the graph of the binomial distribution is bell-shaped, as it must converge to the normal distribution as $N \to \infty$, by the Central Limit Theorem. The fact that Figure 1 is U-shaped suggests that consumers have heterogeneous purchase probabilities and the distribution of purchase probabilities is instead a mixture of binomials.

Figure 1: Distribution of shopping occasions over 12 months



Notes: This figure presents the distribution of consumers purchase occasions over 12 months. By construction, consumers must visit the store at least once in order to be included in the sample, therefore, the distribution is truncated from the left at 1.

Table 2: Estimation of the (marginal) distribution of ρ_i

Class	Pr(Class = i)		ρ_i	
	Estimate	Std. err.	Estimate	Std. err.
Occasional shoppers	0.606	(0.005)	0.162	(0.001)
Regular shoppers	0.240	(0.004)	0.560	(0.005)
Frequent shoppers	0.154	(0.003)	0.936	(0.002)
Observations	17,756			

Notes: This table presents the estimates of a finite mixture generalized linear model with three latent classes, a Binomial(12, ρ_i) family, and a logit link. Robust standard errors, reported in parenthesis, are computed using the delta method.

We approximate this heterogeneity using a finite mixture generalized linear model to obtain an adequate support for the potential values of ρ_i. We fit the model to the data on cumulative purchase occasions over 12 months, using three latent classes, a Binomial(12, ρ_i) family, and a logit link. We do not include demographics in the specification of ρ_i as they are not readily available in the data.

The results are available in Table 2. We recover three broad types of consumers. "Frequent shoppers" have a probability of getting a shopping occasion of 0.94 and constitute 15% of the sample. "Regular shoppers" have a probability of getting a shopping occasion of 0.56

and constitute 24% of the sample. Finally, "occasional shoppers" have a 0.16 probability to visit the store in a given month and constitute 61% of the sample.

Estimation of support. We now turn our attention to estimating the support of the preference parameters α_i^k and β_i^k. Following the suggestion in Fox et al. (2016), we estimate the following likelihood equation,

$$
\begin{aligned}
&\max_{\boldsymbol{a}^k,\, \boldsymbol{b}^k,\, \{\boldsymbol{\delta}_t\}_{t=1,\dots T}} \sum_{i=1}^{M} \ln\left(\prod_{t=1}^{T}\prod_{j=0}^{J} \mathbb{P}_{jt}(\alpha_i^k, \beta_i^k, \boldsymbol{\delta}_t^k)^{Y_{ijt}^k}\right), \\
&\text{s.t.} \quad s_{jt}^k = \overline{\rho}\cdot\frac{1}{M}\sum_{i=1}^{M}\frac{\exp(\delta_{jt}^k+\beta_i^k+\alpha_i^k p_{jt}^k)}{1+\sum_{j'}\exp(\delta_{j't}^k+\beta_i^k+\alpha_i^k p_{j't}^k)},
\end{aligned}
\tag{12}
$$

where all consumers are equally weighted, ρ_i is assumed to be independent of α_i and β_i, and $\overline{\rho}$ is the average probability of obtaining a purchase occasion in the population, computed from Table 2. We show in Appendix C that equation (12) is a special case of the problem described in equation (11) when ρ_i is independent of α_i and β_i.

We assume that

$$
\begin{aligned}
\alpha_i^k &= a_1^k + a_2^k\,\eta_{i1}, \\
\beta_i^k &= b_1^k + b_2^k\,\eta_{i2},
\end{aligned}
\tag{13}
$$

where (η_{i1}, η_{i2}) are independent draws from a Uniform(0, 1) distribution, and $(a_1^k, a_2^k, b_1^k, b_2^k)$ are the parameters to be estimated for each category. Table A.1 provides the estimated parameters and the corresponding supports for the preference parameters.

Estimation on a fixed grid. We define a grid of preference parameters based on the estimated support of α_i^k, β_i^k, and ρ_i. The grid is constructed by interacting the three values of ρ_i in Table 2 with 20 points drawn randomly from the joint support of (α_i^k, β_i^k), leaving us with an initial grid of 60 consumer types. We then estimate the optimization problem described in equation (11), to recover type probabilities $\{\phi_d^k\}_{d=1,\dots,D}$ and the mean utilities $\{\boldsymbol{\delta}_t^k\}_{t=1,\dots,T}$, for each category separately and taking α_d^k, β_d^k, and ρ_d as given.

During the optimization, we restrict the unconditional distribution of ρ_d to match the estimates from Table 2. Therefore, the unconditional distribution of ρ_d is the same across categories. Types with an estimated probability mass below 0.01 are removed iteratively and the estimation is performed again, until all remaining types survive the iterative procedure.[18] Additional details are provided in Appendix C.

4.4 Results

Table 3 presents the estimation results for two randomly selected categories. The randomly selected categories are "yogurt" and "crackers". Detailed results on all categories are available in the Appendix Table A.2 to A.5. Table 3 reports the final fixed grid $\{\alpha_d, \beta_d, \rho_d\}_{d=1,\ldots,D}$, obtained as part of the estimation routine for both categories, and the estimated type probabilities $\{\phi_d\}_{d=1,\ldots,D}$. We note that "yogurt" exhibits 2 consumer types based on preferences, interacted with the shopping occasion probabilities to form 6 consumer types. For "crackers", we identify 5 types based on consumer preferences, and the interaction with the shopping occasion probabilities leads to a total of 9 consumer types (not every combination of (α_d, β_d) and ρ_d survive the iterative procedure). We uncover varying levels of heterogeneity in consumers' preference parameters. For example, there is a large amount of variation in price sensitivities and little variation in the consumer-specific taste parameter in the yogurt category. In contrast, we find more heterogeneity in β_d in the crackers category. Therefore, our estimation yields very different patterns of heterogeneity across categories, both in terms of the number of consumer types and their associated preference parameters.

Table 3: Distribution of consumer types, for selected categories

Dep: Dairy
Cat: Yogurt

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.528	-1.002	0.162	0.494
2	-1.528	-1.002	0.560	0.184
3	-1.528	-1.002	0.936	0.096
4	-0.907	-1.195	0.162	0.112
5	-0.907	-1.195	0.560	0.056
6	-0.907	-1.195	0.936	0.058

Dep: Grocery
Cat: Crackers

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.805	-0.644	0.162	0.542
2	-0.805	-0.644	0.560	0.186
3	-0.805	-0.644	0.936	0.067
4	-0.623	-0.056	0.560	0.013
5	-0.623	-0.056	0.936	0.068
6	-0.435	-0.865	0.162	0.053
7	-0.435	-0.865	0.560	0.041
8	-0.385	-0.097	0.936	0.018
9	-0.266	-0.798	0.162	0.011

Notes: This table presents the distribution of consumer's types, for two randomly selected categories. The $(\alpha_d, \beta_d, \rho_d)$ are the fixed grid points and ϕ_d is the estimated probability to belong to each type.

[18] In practice, removing types with no mass does not change the results meaningfully and it simplifies the evaluation of counterfactual experiments.

Table 4: Consumers' preferences

Category	# Types	α_d^k		β_d^k		ρ_d		δ_{jt}^k	
	(D_k)	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Alcohol									
Beer	7	-0.641	0.134	-0.344	0.108	0.377	0.290	5.493	1.273
Wine	6	-0.516	0.149	0.106	0.232	0.377	0.290	-0.140	1.105
Dairy									
Butter/margarine/spreads	8	-1.126	0.266	-1.302	0.261	0.377	0.290	2.242	1.157
Cheese	11	-0.785	0.105	-0.880	0.102	0.377	0.290	3.719	1.987
Creams/creamers	9	-1.616	0.262	-0.754	0.345	0.377	0.290	4.202	1.857
Fresh eggs	13	-1.243	0.169	-1.351	0.195	0.377	0.290	5.011	0.542
Refrigerated Juice/beverage	12	-0.774	0.140	-0.897	0.181	0.377	0.290	2.447	0.773
Milk	13	-1.593	0.132	-1.702	0.111	0.377	0.290	10.748	3.750
Yogurt	6	-1.388	0.259	-1.045	0.081	0.377	0.290	4.450	1.642
Frozen food									
Ice cream	7	-0.717	0.071	-0.714	0.187	0.377	0.290	2.360	1.924
Frozen pizza	8	-1.411	0.227	-1.321	0.027	0.377	0.290	7.141	1.926
Frozen potatoes/onions	7	-1.350	0.641	-0.664	0.083	0.377	0.290	0.100	2.051
Frozen vegetables	10	-1.327	0.126	-1.411	0.113	0.377	0.290	8.343	3.865
Grocery									
Cereal	9	-0.523	0.186	-0.648	0.139	0.377	0.290	-1.178	1.258
Coffee	6	-0.907	0.143	-0.982	0.030	0.377	0.290	3.902	2.183
Crackers	9	-0.742	0.134	-0.608	0.193	0.377	0.290	-0.580	1.784
Snacks grocery	9	-1.134	0.052	-1.202	0.085	0.377	0.290	8.337	5.072
Soft drinks/mixers	12	-1.396	0.067	-1.398	0.094	0.377	0.290	15.256	8.514
Water	8	-0.402	0.099	-0.392	0.155	0.377	0.290	-0.502	0.586
Package meat									
Bacon	8	-1.381	0.119	-1.392	0.049	0.377	0.290	10.247	2.536
Dinner sausage	12	-1.108	0.120	-1.155	0.102	0.377	0.290	8.998	2.664
Taxable grocery									
Bleach/stain removers	9	-1.542	0.655	-0.810	0.140	0.377	0.290	-1.117	1.330
Laundry detergent	6	-0.912	0.195	-0.575	0.038	0.377	0.290	1.305	1.892
Pet food	6	-1.179	0.224	-1.458	0.050	0.377	0.290	5.615	2.123

Notes: This table presents a summary of the estimated consumer preferences. The detailed estimation results are available in Table A.2 to Table A.5. The number of types is the final number of grid points in the $(\alpha_d, \beta_d, \rho_d)$-space. All averages and standard deviations are constructed using the estimated type probabilities $\{\phi_d\}_{d=1,\ldots,D}$. Average and standard deviations over product mean utilities δ_{jt} are unweighted.

We provide a summary of this heterogeneity within and across categories in Table 4. Categories exhibit between 6 and 13 consumer types, with an average of 8.8 types (the median is 8.5). These are the grid points that survive our iterative procedure described above. We report for each category the average price sensitivity α_d^k, the average consumer-specific taste parameter β_d^k, the average purchase occasion probability ρ_d, and the average mean utility of products δ_{jt}^k. These statistics are computed using the fixed grid and the estimated type probabilities when applicable (the statistics are unweighted in the case of product-specific mean utilities). The corresponding standard deviations are also reported. They provide a measure of within-category heterogeneity for each parameter α_d, β_d and ρ_d, or a measure of product differentiation for δ_{jt}, respectively.

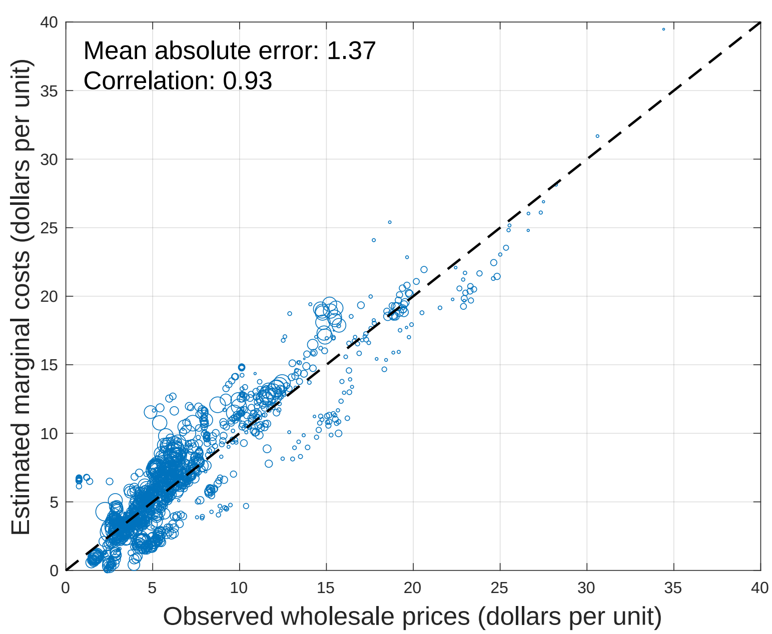
By construction, the average and standard deviation of the purchase occasion probability is constant across categories, since we restrict the estimation to match the unconditional distribution in Table 2. For completeness, we also report the marginal distributions of consumer preferences (α_d, β_d), obtained by integrating over ρ_d. The results are available in Figure A.1 and A.2.

4.5 Marginal costs

We recover the marginal cost of each product by using the set of first-order conditions for the uniform price case, see equation (2). These reflect the price the supermarket pays to the suppliers and any additional retail costs. In counterfactual experiments, these estimated marginal costs are used to compute personalized prices, so they are crucial to the analysis.

Our data include wholesale prices at the product level for all observed categories and products. We compare our estimated marginal costs against the observed wholesale prices to assess the model's goodness of fit. Figure 2 presents a scatter plot of the estimated marginal costs and the observed wholesale price. Each data point is represented by a circle of size proportional

Figure 2: Estimated marginal costs vs observed wholesale prices



Notes: Observed wholesale prices and estimated marginal costs from our estimation procedure. Each circle represents one product weighted by its market share.

to the product's market share. Overall, our estimates lie close to the 45-degree line. The correlation between the observed and the estimated marginal costs is 0.93 and the (weighted) mean absolute error is $1.37. We take these goodness-of-fit values as a strong sign that our depiction of the market structure is highly accurate.

5 The Impact of Price Personalization

5.1 Setup

We perform counterfactual simulations using our estimated parameter values to study the information acquisition process of our grocery store and different pricing strategies given observed consumers' purchases over time. The implicit assumption is that the chosen grocery store has access to the technology that enables price personalization, and we vary the amount of information the firm has on consumers by restricting the length of consumer histories. We then compare the market outcomes from varying history lengths to two benchmarks: the case where the grocery store has no information and has to set uniform prices, and the case where it has perfect information on consumers and can perfectly personalize prices.[19] Since we estimate discrete consumer types, the perfect discrimination benchmark corresponds to third-degree price discrimination.

For each history of length τ, denoted $\mathbf{h}^\tau$, the information set of the seller includes the set of products available, the price of each product, the set of consumers with a purchase occasion, and consumer choices for each period $t = 1, \ldots, \tau$. The seller uses this information to update its belief about consumers' types and can solve for the vector of personalized prices $\mathbf{p}(\mathbf{h}^\tau)$ using the first-order conditions in equation (7) and the posterior densities constructed from purchase histories $\phi_d(\mathbf{h}^\tau) = f_{\Theta|\mathbf{h}^\tau}(\Theta|\mathbf{h}^\tau)$.

We use the counterfactual simulations to answer three broad sets of questions. First, we want to understand the impact of acquiring consumer data on the distribution of personalized prices. If data on past purchases are informative of consumers' preferences, we would expect the distribution of personalized prices to diverge from uniform pricing towards the perfect discrimination benchmark as we increase the amount of information the seller has access to. Second, we want to characterize the seller's information acquisition process as more consumer

[19] In our two benchmark cases, the counterfactual prices can be obtained from the first-order conditions (2) and (8), and no information on past purchases is required.

data become available. We propose a new measure of information acquisition based on the level of uncertainty the seller has with respect to consumers' preferences and we track its evolution for varying history lengths. We show that more information on consumers correlates with higher profits for the seller. Finally, we use our setup to study the value of consumer data.

5.2 Price dispersion

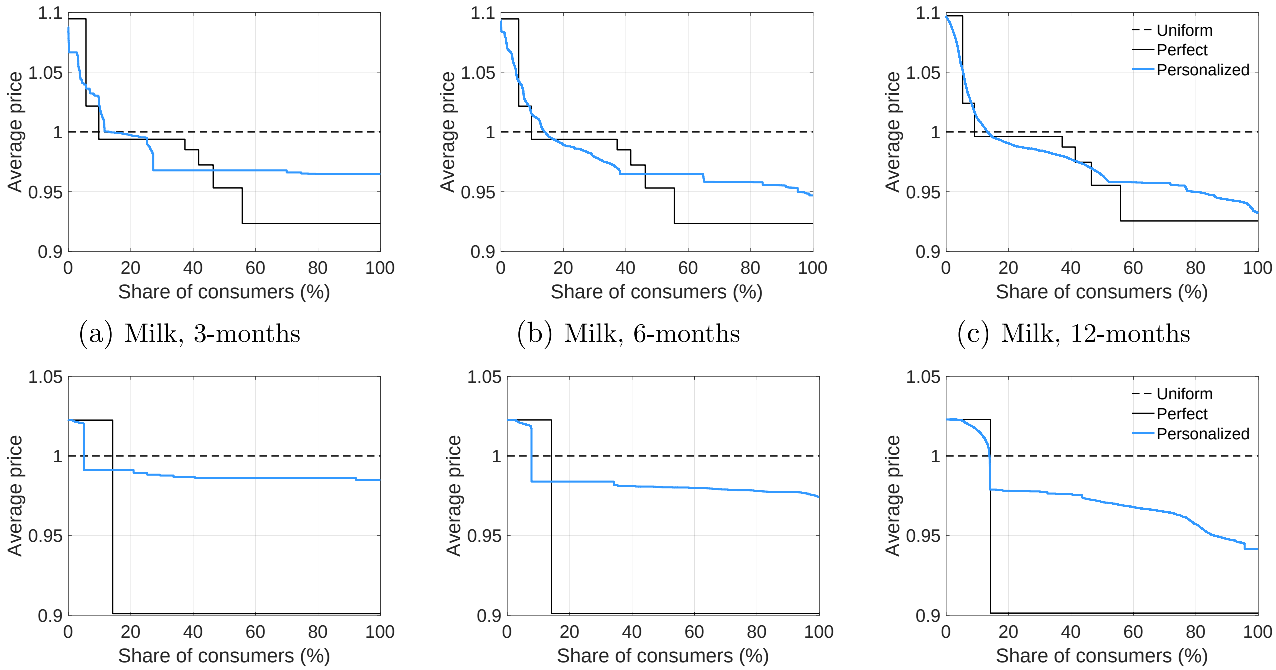
We begin our analysis by studying the pricing behavior of the seller as it acquires more consumer data on past purchases. We evaluate the posterior belief of the seller and the prices it sets for each observed consumer history. We focus on 3-months, 6-months, and 12-months histories.[20] Figure 3 represents the distribution of personalized prices over the distribution of consumers for two randomly selected categories different from those shown in the previous section to showcase the span of our data ("milk" and "coffee"). Prices are averaged over products within each category to form a unique price index. We also provide the distribution of prices from our two benchmarks, and all prices are normalized by the average uniform price. For completeness, we provide additional figures for the omitted categories in Figure A.3 to A.8.

Figure 3 makes evident that the distribution of prices moves toward the perfect discrimination benchmark as more information becomes available to the seller: the area between the personalized price and the perfect discrimination curves is reduced as we increase the length of the histories. Another notable observation is that prices seem to converge faster for the most price-insensitive consumers, which receive supra-uniform prices, than for price-sensitive consumers, which receive infra-uniform prices. This is particularly salient for the "coffee" category, see Figure 3, panels (d), (e), and (f).

On the one hand, price-insensitive consumers are more likely to purchase one of the inside products whenever they obtain a purchase occasion, and the seller quickly realizes through Bayesian updating that repeated purchases cannot be attributed to price-sensitive consumers. This leads the seller to quickly increase prices for consumers which repeatedly purchase

[20]The length of purchase histories in this paper contrasts with studies focusing on pricing algorithms that use A / B-type tests over much longer periods of time, which typically need large amounts of firm-consumer interactions to provide a definitive answer. The choice of shorter purchase histories showcases the advantages of using Bayesian updating. Alternatively, short time spans captures the idea that consumers' preferences may change over time, and the seller may want to focus only on the latest information available, e.g., the last year of data.

Figure 3: Price dispersion, by history length



(a) Milk, 3-months (b) Milk, 6-months (c) Milk, 12-months

(d) Coffee, 3-months (e) Coffee, 6-months (f) Coffee, 12-months

Notes: Distributions of prices over the distribution of consumers for two randomly selected categories, after observing a purchase history of three, six, and twelve months respectively. The selected categories are "milk" and "coffee".

products. On the other hand, choosing the outside good repeatedly can be attributed both to price-sensitive and price-insensitive consumers. In this case the seller gathers less information when consumers do not buy anything (or when they do not obtain a purchase occasion). Convergence to the low prices is slower because there remains a significant probability that consumers who buy nothing belong to the price-insensitive group.

Finally, we note that the vast majority of consumers receive infra-uniform prices as a result of personalization. This suggests two channels for the seller to extract surplus from consumers. On the one hand, it charges high prices to a small group of price-insensitive (and very profitable) consumers. On the other hand, it lowers prices for everyone else, which increases profits through a market expansion effect.

We broaden our perspective and present a summary of the predicted price dispersion at the category level. In what follows, we focus on 12-month length histories. The results are

Table 5: Price dispersion by category, after 12-months history

Category	Surcharge/discount over uniform price (%)							Share
	Mean	Std. dev.	Min.	25th perc.	Median	75th perc.	Max.	infra-uniform
Alcohol								
Beer	-0.21	1.53	-1.89	-1.16	-0.63	0.69	11.83	0.65
Wines	-3.87	2.93	-7.55	-6.45	-5.17	-0.45	0.53	0.93
Dairy								
Butter/margarine/spreads	-4.37	5.44	-13.16	-7.97	-5.73	-0.58	15.62	0.77
Cheese	-2.44	3.60	-8.23	-5.06	-4.05	-0.10	9.02	0.76
Creams/creamers	-3.26	4.06	-9.65	-6.29	-2.73	-0.97	16.23	0.91
Fresh eggs	-3.40	4.55	-9.08	-6.39	-5.27	-0.76	15.97	0.78
Refrigerated juice/beverage	-5.79	6.48	-12.75	-10.71	-7.54	-3.79	15.04	0.80
Milk	-2.51	3.32	-6.80	-4.49	-3.62	-1.32	9.70	0.87
Yogurt	-1.45	1.50	-3.97	-2.31	-1.48	-1.08	1.49	0.84
Frozen food								
Ice cream	-3.13	3.40	-5.90	-4.80	-4.51	-4.37	14.77	0.84
Frozen pizza	-0.51	0.66	-1.34	-1.13	-0.42	-0.05	3.44	0.87
Frozen potatoes/onions	-4.30	4.02	-5.31	-5.22	-5.18	-5.12	22.65	0.96
Frozen vegetables	-1.03	1.01	-2.20	-1.72	-1.31	-0.80	2.87	0.84
Grocery								
Cereal	-10.58	17.65	-25.22	-20.59	-16.49	-15.11	119.74	0.80
Coffee	-2.66	2.11	-5.84	-3.73	-2.89	-2.21	2.29	0.86
Crackers	-5.23	9.12	-11.25	-9.80	-9.01	-5.31	56.03	0.83
Snacks grocery	-1.07	1.42	-2.52	-2.07	-1.91	-0.41	2.68	0.81
Soft drinks/mixers	-0.72	1.75	-3.36	-1.90	-1.45	0.34	5.96	0.69
Water	-11.43	17.69	-24.71	-21.98	-19.09	-7.39	88.40	0.80
Packaged meat								
Bacon	-1.29	1.06	-2.43	-1.95	-1.47	-1.02	4.39	0.92
Dinner sausage	-1.93	1.82	-4.51	-3.16	-2.33	-1.29	5.65	0.87
Taxable grocery								
Bleach/stain removers	-8.44	10.09	-25.04	-16.47	-5.02	-1.13	36.25	0.95
Laundry detergent	-0.83	1.19	-2.96	-1.88	-0.50	0.08	3.71	0.63
Pet food	-0.01	0.01	-0.03	-0.01	-0.01	-0.00	0.00	0.93

Notes: This table presents the distribution of prices, by categories, after observing consumers for 12 months. All prices are averaged over products within a category. "Surcharge/discount over uniform price" represents the difference between personalized prices and uniform prices, in percentage. Negative values indicate a discount and positive values a surcharge. "Share infra-uniform" reports the share of consumers that obtain a lower price under personalization compared to the uniform price in each product category.

presented in Table 5.

We uncover significant heterogeneity across categories: price personalization based on past purchases can lead to varying levels of price dispersion for different food categories. For example, we find that uniform prices are essentially optimal in the case of "pet food". If price sensitivity correlates with pet ownership, there is no point in lowering prices for consumers who do not have pets, as they would never buy pet food. In this case, there is no room for a market expansion effect and the seller simply sets a uniform price and ignores any information on past purchases.

In other cases, we find that the level of discounts and surcharges offered to consumers can be large, from -25% to +120% of the list price in the most extreme categories (e.g., "cereal" and "water"). In most categories, the discounts offered or the surcharges are in the range of a few percentage points. This variation in price dispersion across categories can be explained by the underlying distribution of consumers' preferences, at the category level, and how much information is revealed by purchase histories.

On average, consumers pay slightly lower prices when the seller personalizes prices, in the range of a few percentage points. Similarly to our findings on "milk" and "coffee", we find that the vast majority of consumers receive infra-uniform prices in all categories.

5.3 Information acquisition

One of the main goals of this paper is to characterize the learning process of the seller as it acquires more information on consumers' past purchases. To that end, we propose a simple index that characterizes information acquisition from purchase histories.

Consider a preference parameter $\theta_i \in \{\alpha_i, \beta_i\}$ and a history of length τ, where we omit the category index k to ease readability. Our chosen measure of information is based on the conditional variance of θ_i given $\mathbf{h}^\tau$, normalized by the unconditional variance of θ_i. The ratio of these variances measures the degree of uncertainty that the seller faces compared to the no information case.

Formally, our information index is defined as

$$\mathcal{I}_\theta(\tau) = 1 - \left(\frac{\sigma^2_{\theta|\mathbf{h}^\tau}}{\sigma^2_\theta}\right)$$

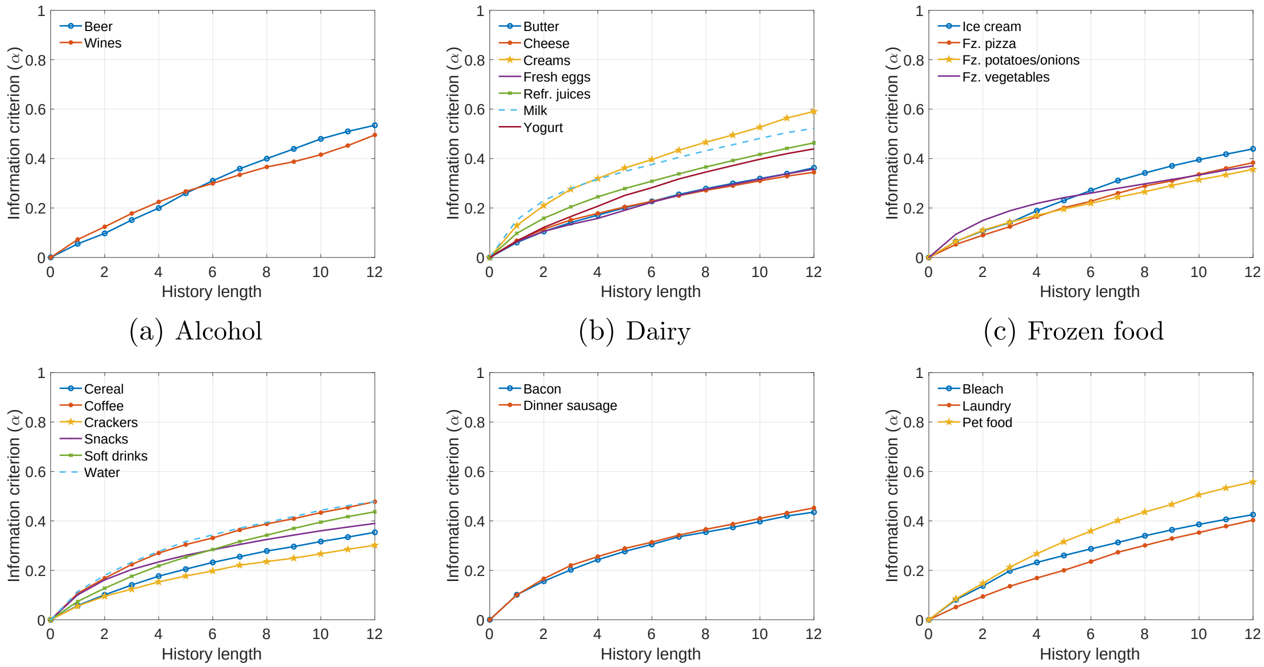
where

$$\sigma^2_{\theta|\mathbf{h}^\tau} = \sum_{\ell=1}^{L}\left(\sum_{d=1}^{D_k}\big(\theta_d - \mathrm{E}(\theta_i|\mathbf{h}^\tau_\ell)\big)^2 \cdot \phi_d(\mathbf{h}^\tau_\ell)\right)\cdot\left(\frac{M_{\mathbf{h}^\tau_\ell}}{M}\right), \text{ and}$$

$$\sigma^2_\theta = \left(\sum_{d=1}^{D_k}\big(\theta_d - \mathrm{E}(\theta_i)\big)^2 \cdot \phi_d\right).$$

Recall that $\ell = 1, ..., L$ indexes separate histories, $M_{\mathbf{h}^\tau_\ell}$ is the number of consumers with history $\mathbf{h}^\tau_\ell$, and $\phi_d(\mathbf{h}^\tau_\ell)$ is the probability that a consumer belongs to type d conditional on observing history $\mathbf{h}^\tau_\ell$. Our information index, $\mathcal{I}_\theta(\tau)$, takes a value of 0 when no information on purchase histories is available and a value of 1 when the firm has perfect information on consumers.[21] If information on purchase histories is informative of consumers' types, the information criterion should be increasing in history length and converge towards 1.

Figure 4: Information acquisition (α_i)



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

(d) Grocery (e) Packaged meat (f) Taxable grocery

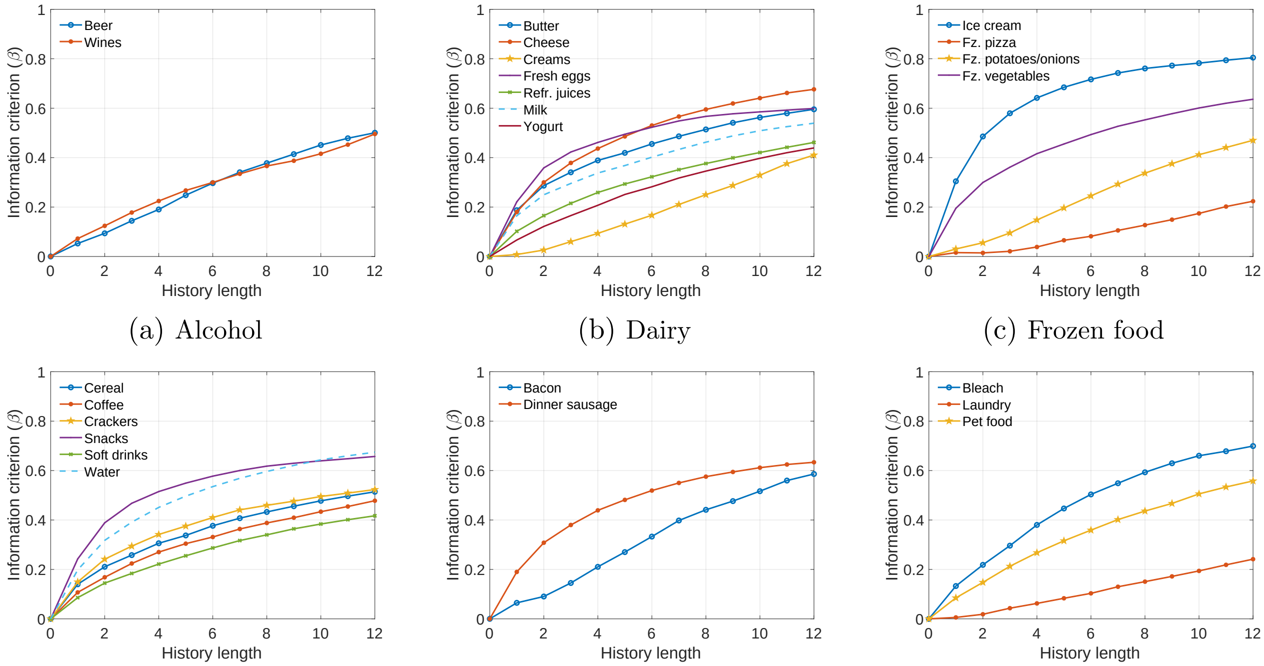
Notes: In each panel, the lines represent the information acquisition index for the price sensitivity α_i, for goods in six different product categories as a function of the history length.

We compute our information index for each preference parameter $\theta_i^k \in \{\alpha_i^k, \beta_i^k\}$ for each product category k as a function of the length of the histories available to the seller. The results for the price sensitivity parameter α_i^k are presented in Figure 4 and the results for the taste parameter β_i^k are presented in Figure 5.

A few observations are worth noting. First, consumer purchase histories are informative about consumer types: $\mathcal{I}_\theta(\tau)$ is strictly increasing in τ for all categories. After observing consumers for 12 periods, the seller is able to reduce the variance of consumers' price sensitivity by 30 to 60%, depending on the product category. For the consumer taste parameter β_i, the variance is instead reduced by 25% to 80% after 12 periods. Although those ranges are relatively

[21]When the seller has perfect information, $\phi_d(\mathbf{h}_\ell^\infty)$ becomes degenerate and $\sigma^2_{\theta|\mathbf{h}^\infty} \to 0$. Therefore, $\mathcal{I}_\theta(\infty) \to 1$.

Figure 5: Information acquisition (β_i)



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

(d) Grocery (e) Packaged meat (f) Taxable grocery

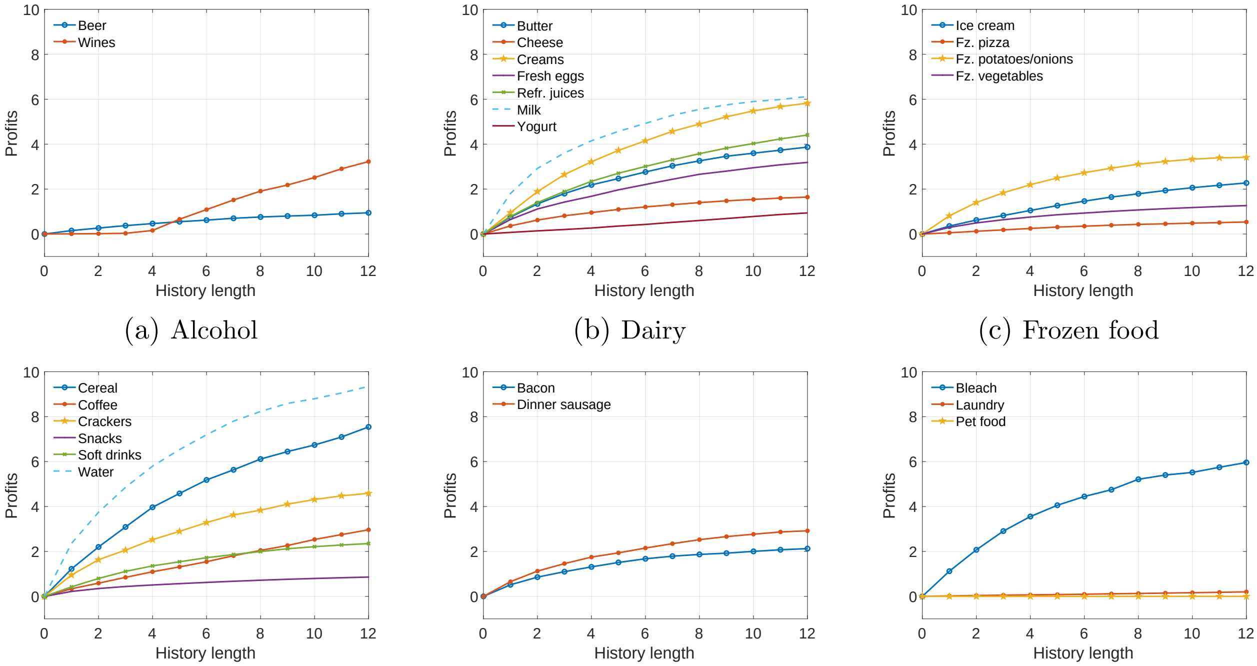
Notes: In each panel, the lines represent the information acquisition index for the intercept β_i, for goods in six different product categories as a function of the history length.

large, these findings show that large reductions in uncertainty about consumers' types can be achieved with relatively few interactions with consumers.

Second, we note that the returns on information are not necessarily decreasing. For example, information seems to have increasing returns with respect to α_i for the "wine" category over at least part of the range. We find evidence of increasing returns with respect to β_i for more categories for at least some part of the history length: "wine", "cream", "frozen potatoes", "frozen vegetables", "bacon", "laundry detergent", and "pet food".

Whether or not information on past purchases is profitable for the seller depends on several crucial factors, for example, the distribution of the preferences of the underlying consumers. As seen in the previous section, the seller is able to personalize prices more accurately for price-insensitive consumers who repeatedly purchase products. On the other hand, the seller struggles to pin down the type of consumers that almost never purchase anything, and this contributes to the large aggregate uncertainty about consumers' types.

Figure 6: Impact of personalized pricing on aggregate profits (% from uniform)



(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 from personalized pricing over uniform pricing, in percentage.

5.4 Profitability of personalized pricing

A natural question is to what extent personalized pricing increases profits. We compute the expected change in profits, from uniform to personalized prices, using histories of varying length. The results are presented in Figure 6.

We find that, unsurprisingly, the seller's expected profits increase in each category as more information on consumers' past purchases becomes available. In general, profits increase at a decreasing rate with few exceptions (e.g., the "wine" category is the most obvious example) and the seller's profits can increase by up to 9%. This is economically significant considering that we are relying on very few time observations.

For other categories, we find that additional information about consumers' purchases leads to negligible increases in profits, despite the seller reducing its uncertainty about consumers' types. This is most salient for the "laundry detergent" and "pet food" categories: the seller reduces its uncertainty about consumers' price sensitivity by 40% and 60% respectively, yet, profits do not change. Our interpretation is that this pattern emerges in categories where there is no scope for a market expansion effect, due to the underlying preferences of consumers. In this case, price discrimination becomes less relevant as the seller cannot induce the price-sensitive consumers to buy the product by offering a better price.

To shed some light on the value of personalized pricing to the seller, we compute the change in margin per unit sold at uniform prices that is equivalent to the gains from personalization. The results are presented in Table 6.

We report on the total quantity sold, the average price paid by consumers, and the average margin per unit under uniform pricing. Then, keeping prices and quantities fixed to the uniform price levels, we compute by how much margins need to increase to match profits from price personalization based on past purchases or perfect discrimination.

Table 6: Change in margin under uniform pricing to attain discriminatory profits

Category	Uniform pricing outcomes			ΔMargin equivalent to price personalization (%)			
	Quantity	Unit price	Unit margin	3-month hist.	6-month hist.	12-month hist.	Perfect
Alcohol							
Beer	169	21.24	3.47	0.6	1.0	1.5	4.2
Wines	181	25.44	11.14	0.0	1.6	4.9	13.0
Dairy							
Butter/margarine/spreads	997	4.90	1.72	2.6	4.0	5.4	10.6
Cheese	1,142	7.22	2.07	1.3	1.9	2.5	5.3
Creams/creamers	404	5.90	1.33	4.2	6.5	9.0	15.1
Fresh eggs	1,266	5.05	1.45	2.1	3.2	4.6	9.1
Refrigerated juice/beverage	824	7.21	2.40	2.8	4.4	6.3	11.5
Milk	1,781	8.20	1.39	5.4	7.4	8.9	13.4
Yogurt	266	6.78	1.25	0.3	0.6	1.4	3.9
Frozen food							
Ice cream	425	7.71	1.91	1.2	2.2	3.3	6.2
Frozen pizza	178	9.20	1.20	0.3	0.5	0.8	2.1
Frozen potatoes/onions	473	5.38	2.05	2.6	3.9	4.7	6.7
Frozen vegetables	726	9.93	1.15	0.9	1.3	1.7	3.8
Grocery							
Cereal	567	6.00	3.92	4.5	7.6	10.8	23.6
Coffee	306	11.11	2.17	1.2	2.2	4.1	8.1
Crackers	653	4.56	2.02	3.1	5.0	6.8	14.2
Snacks grocery	1,183	8.52	1.32	0.7	0.9	1.3	2.6
Soft drinks/mixers	1,518	10.75	1.30	1.8	2.8	3.7	6.8
Water	855	7.66	4.71	6.6	9.6	12.0	20.6
Packaged meat							
Bacon	542	10.24	1.10	1.5	2.4	2.9	4.8
Dinner sausage	909	11.79	1.61	2.1	3.0	4.0	7.2
Taxable grocery							
Bleach/stain removers	241	4.70	2.86	3.9	5.9	7.7	20.9
Laundry detergent	112	9.64	2.40	0.1	0.1	0.3	2.4
Pet food	120	15.10	2.16	0.0	0.0	0.0	0.0

Notes: This table represents the cost reduction that is required to reach discriminatory profits given uniform prices. All cost reductions are in percentage points. The average unit price and average marginal costs are computed using the data from December 2018 only.

We present the results for personalized prices based on 3-, 6-, and 12-month histories, and for the perfect discrimination case. After observing consumers for 12 periods, the seller's increase in profits is equivalent to an increase in the per unit margin of 0 to 12%, with an average over all categories of 4.5%.

We compare these results with the perfect discrimination outcomes. We find, with perfect information on consumers, the seller's increase in profits is equivalent to an increase in the per unit margin between 0 and 23.6%, with an average of 9.4%. This implies that the seller can capture around half the gains from perfect discrimination with only 12 interactions with consumers.

There is, however, large variation across categories. As explained previously, a small number of categories are very profitable for the seller while others provide essentially the same profits with or without personalization. To understand what drives these differences at the category level, we study the determinants of the profitability of personalized pricing through a descriptive regression analysis. The results are presented in Table 7.

We split the data into four broad categories: information (e.g., our information indices), product characteristics, consumer characteristics, and market coverage. The dependent variable is the change in profits, in percentage points, from uniform pricing to price personalization based on purchase histories of different lengths. We stack the data from all categories and different history lengths.

We perform a separate regression for each regressor separately, then for all regressors together, and we focus on the within R-squared value, which informs us about the importance of each regressor in explaining profits.

We begin with our information indices. Both are positively correlated with profits (after accounting for history length fixed effects) which indicates that reducing uncertainty about consumers' types leads to higher profits. The R-squared are 0.03 and 0.051, which means that these variables account for roughly 3% and 5.1% of the gain in profits only. This suggests that other features of the market at the category level could explain the differences in profitability.

We consider three product characteristics and five consumer characteristics. The product characteristics are the number of products, the average quality of products (i.e., the average quality index, δ_j), and the standard deviation of product quality (i.e., a measure of the

Table 7: Determinants of the profitability of personalized pricing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Information indices:												
$\mathcal{I}_\alpha(\tau)$	0.057**											0.060***
	(0.024)											(0.017)
$\mathcal{I}_\beta(\tau)$		0.030***										0.006
		(0.008)										(0.005)
Product characteristics:												
# Products			0.004***									0.008***
			(0.001)									(0.001)
$E(\delta_j)$				-0.001								-0.005***
				(0.000)								(0.001)
$SD(\delta_j)$					-0.002***							0.001
					(0.000)							(0.001)
Consumer characteristics:												
# Types						0.002***						0.002***
						(0.000)						(0.001)
$E(\alpha_i)$							0.006					-0.005
							(0.004)					(0.007)
$SD(\alpha_i)$								0.020***				0.057***
								(0.005)				(0.008)
$E(\beta_i)$									0.007**			0.014**
									(0.003)			(0.006)
$SD(\beta_i)$										0.103***		0.028*
										(0.009)		(0.015)
Market coverage:												
Coverage											0.001***	0.003***
											(0.000)	(0.000)
Within R-squared	.03	.051	.069	.007	.041	.094	.015	.028	.029	.188	.077	.608
Observations	288	288	288	288	288	288	288	288	288	288	288	288

Notes: The dependent variable the change in profits between uniform and personalized pricing, $\pi(\mathbf{p}(\mathbf{h}^\tau))/\pi(\mathbf{p}^U) - 1$. Variable τ represents the history length. The dataset was assembled by stacking the results from the counterfactual simulations on 24 food categories and 12 different history lengths. The information indices $\mathcal{I}_\alpha(\tau)$ and $\mathcal{I}_\beta(\tau)$ are computed as described in Section 5.3. Variable α_i represents the price sensitivity, β_i is the consumer's preference for the inside good, and δ_j represents product quality. The product quality is normalized such that the lowest quality in each category is equal to zero. $E(\cdot)$ is the expectation operator and $SD(\cdot)$ is the standard deviation. "Coverage" is the total market share of the inside goods. All regressions include history length fixed effects. Robust standard error in parenthesis. Significance: * $<$ 0.10; ** $<$ 0.05; *** $<$ 0.01.

degree of product differentiation in a category). The consumer characteristics include the number of consumer types, the average price sensitivity, the standard deviation of the price sensitivity, the average consumer-specific taste parameter, and the standard deviation of the consumer-specific taste parameter. These consumer characteristics aim at capturing average characteristics and the degree of heterogeneity in consumer characteristics.

We find that the number of products, the number of consumer types and the standard deviation of product quality explain 6.9%, 9.4%, and 4.1% of the variation in profits across categories respectively. Perhaps surprisingly, the standard deviation of β_i is by far the strongest explanatory variable, explaining 18.8% of the variation in profits. The seller therefore increases its profits more in markets where consumers have very heterogeneous tastes for the inside good, captured by the parameter β_i.

Finally, we consider market coverage, defined by the market share of the inside good, as an explanatory variable. Market coverage is potentially important. There is more room for market expansion in markets with low coverage. At the same time, markets with high coverage may have a larger share of price-insensitive consumers from which more profits can be extracted. Thus, the direction of the effect, and the explanatory power of market coverage are informative about the channel which leads to higher profits.

We find that market coverage positively correlates with profits, suggesting that the market expansion channel is not as profitable as extracting profits from current consumers. The within R-squared is 0.077. When considering all regressors together, we obtain a within R-squared of 0.608, meaning that we explain roughly 61% of the increase in profits realized from price personalization with these covariates.

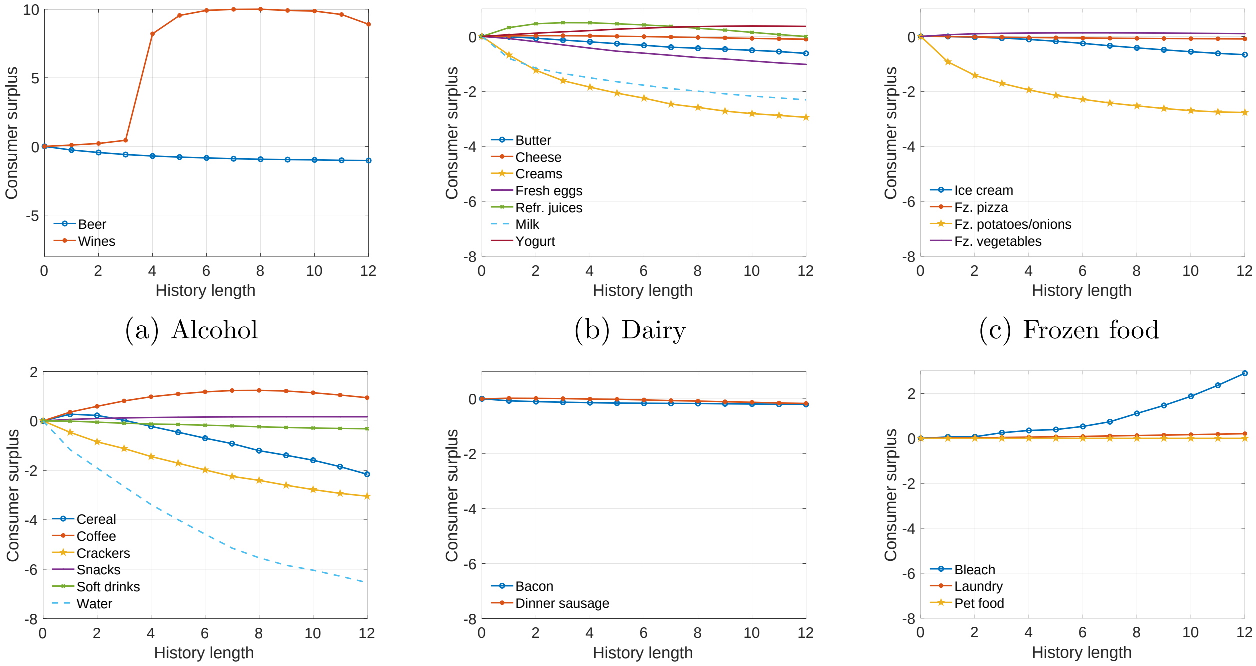
5.5 Impact of personalization on consumers

We finally turn our attention to the impact of price personalization on consumers. We consider the change in aggregate consumer surplus, in percentage points, from uniform to personalized pricing based on purchase histories of varying lengths. The results are presented in Figure 7.

There are several notable observations. Price personalization can lead to gains in consumer surplus, in some categories. Rhodes and Zhou (2022) show that under specific circumstances, personalized pricing can increase consumer surplus, if for example it improves market coverage enough. In practice, whether or not price personalization harms consumers in the aggregate is an empirical question. In our case, consumer surplus rises for several categories, e.g., “wine”, “refrigerated juices”, “yogurt”, “frozen pizza”, “frozen vegetables”, “coffee”, “snacks”, “laundry detergent”, and “pet food”.

For all other categories, consumer surplus decreases and the loss can reach up to 6.5% compared to the uniform pricing benchmark. There are two contributing factors to the decline in consumer surplus. 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.

Figure 7: Impact of personalized pricing on aggregate consumer surplus (% from uniform)



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

(d) Grocery (e) Packaged meat (f) Taxable grocery

Notes: Each line represents the evolution the change in aggregate consumer surplus, from personalized pricing over uniform pricing, in percentage.

Finally, the impact of personalization on consumer surplus is not monotonic. This is especially true for the categories where consumer surplus improves as a result of personalization: we see a clear inverted U-shape, suggesting that with more information, consumer surplus would eventually decline.

These aggregate results hide significant heterogeneity at the consumer level. Table 8 provides the distribution of consumer surplus, by category, for 12-month purchase histories.

To preserve comparability across categories, we normalize the change in consumer surplus by the average consumer expenditure in each category. As shown in Table 5, the vast majority of consumers (over 80%) receive infra-uniform prices under price discrimination. For these consumers, consumer surplus can increase by up to 18.5% of the average expenditure by category. For most of the consumers with infra-uniform prices, however, the gains in consumer surplus are small, totalizing only a few percentage points of the average expenditure.

A smaller group of consumers receive supra-uniform prices and experience losses in terms

Table 8: Distributional impact of personalized pricing on consumers (% of avg. expenditure)

Category	ΔConsumer surplus (% of avg. expenditure)						
	Mean	Std. dev.	Min.	10th pct.	Median	90th pct.	Max.
Alcohol							
Beer	-0.05	1.48	-20.97	-0.10	0.05	0.17	2.70
Wines	1.31	1.77	-3.90	0.05	0.99	4.32	4.50
Dairy							
Butter/margarine/spreads	-0.08	6.07	-37.62	-1.71	0.67	3.37	8.01
Cheese	-0.02	4.80	-26.03	-3.74	0.87	3.65	7.91
Creams/creamers	-0.13	4.43	-43.76	0.04	0.14	1.16	4.81
Fresh eggs	-0.13	5.53	-43.84	-2.55	0.90	3.47	6.95
Refrigerated juice/beverage	0.00	7.33	-49.25	-3.03	1.08	4.56	7.61
Milk	-0.36	9.65	-53.04	-2.99	1.43	5.20	9.92
Yogurt	0.01	0.44	-1.64	-0.47	0.06	0.36	0.56
Frozen food							
Ice cream	-0.04	2.40	-14.17	-1.00	0.37	1.40	2.47
Frozen pizza	0.00	0.55	-7.94	0.00	0.02	0.17	0.45
Frozen potatoes/onions	-0.20	5.47	-44.13	0.21	0.23	2.15	3.53
Frozen vegetables	0.01	1.69	-11.59	-0.38	0.24	0.99	1.75
Grocery							
Cereal	-0.44	15.00	-134.76	-5.52	1.73	6.82	11.69
Coffee	0.05	1.74	-8.03	-1.18	0.26	1.52	1.81
Crackers	-0.35	7.64	-74.75	-1.80	0.74	2.80	4.72
Snacks grocery	0.02	2.01	-7.34	-2.54	0.54	1.60	3.54
Soft drinks/mixers	-0.04	4.58	-36.16	-2.29	0.79	3.42	7.25
Water	-2.16	26.10	-260.58	-16.23	2.98	11.09	18.54
Packaged meat							
Bacon	-0.01	2.17	-19.71	0.05	0.18	0.71	1.10
Dinner sausage	-0.02	4.63	-36.99	-0.60	0.57	2.13	3.79
Taxable grocery							
Bleach/stain removers	0.15	6.02	-61.98	0.02	0.15	3.05	5.98
Laundry detergent	0.00	0.12	-0.62	-0.02	-0.01	0.13	0.15
Pet food	0.00	0.00	-0.01	0.00	0.00	0.00	0.00
Observations	17,756						

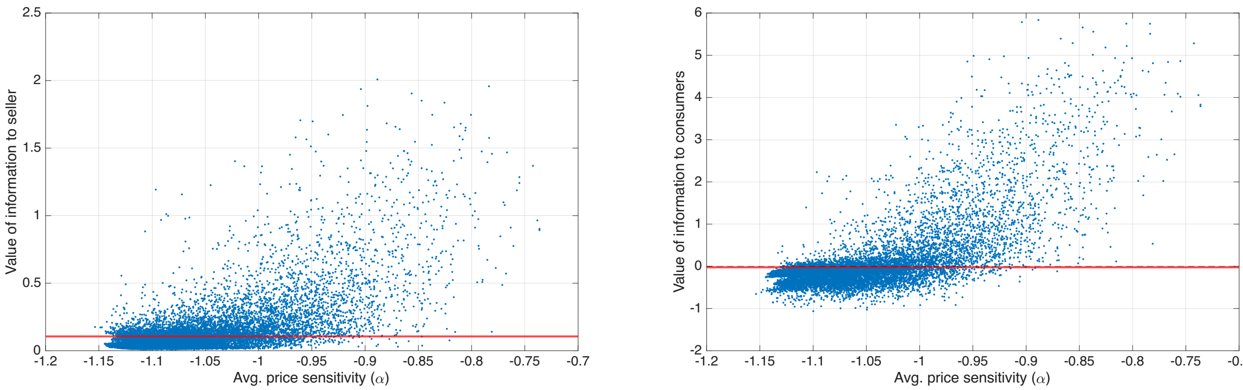
Notes: This table presents the distribution of the change in consumer surplus between personalized prices and uniform prices, as a percentage of the average consumer expenditure per category. Personalized prices are based on 12-month histories.

of consumer surplus. These losses are much larger than the gains from infra-uniform prices, by an order of magnitude. Therefore, the impact of personalization on consumers is mostly redistributive, shifting surplus from price-insensitive consumers to price sensitive consumers. Whether or not surplus increases overall depends on how large losses to few consumers balance out with small gains spread over many consumers.

6 The Value of Information

6.1 Distributional effects

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



(a) Willingness-to-pay to acquire data

(b) Compensation required to share data

Notes: These figures depict the seller's willingness-to-pay to acquire each consumer's purchase history and the compensation that is required to offset the harm done to consumers from personalized pricing. The willingness-to-pay is the change in profits from uniform to personalized pricing and the compensation is (minus) the change in consumer surplus from uniform 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.

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,

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

where the conditional expectation is the expected price sensitivity of consumer i in category k, given its purchase history, and w_k are category weights, common to all individuals, based on the category revenues observed in the data. Then, for each individual history, we plot the average price sensitivity of the consumer against the change in seller profits and the change in consumer surplus from personalization. These correspond to the value of information to the seller and to the consumer respectively, in the sense that it is the amount of money that makes the seller or the consumer indifferent between using the consumer data to personalize prices and uniform prices.

The results are presented in Figure 8. Panel (a) depicts the maximum amount the seller is willing to pay to acquire each consumer's history, while panel (b) presents the compensation that is required to make each consumer as well off under personalized pricing, compared to uniform pricing. All values assume a purchase history of 12 months.

We find that the supermarket is able to extract at least some surplus from each and every consumer (the average monthly expenditure is $11.60 per consumer). However, there is a large amount of variation: the retailer is able to extract as much as $2 in extra profits from the most price insensitive consumers, representing around 17% of the average consumer basket. Consumer surplus on the other hand slightly increases for price-sensitive consumers. In this case, price discrimination leads to a win-win scenario, and consumers need not be compensated for providing their personal data. For price-insensitive consumers, consumer surplus decreases and the magnitude is around two to three times as large as the gains in profits. Extracting surplus from these low-sensitivity consumers therefore, leads to wasted total surplus, as the compensation required to make consumers indifferent is larger than the potential gains in profits.

Nevertheless, Figure 8 suggests that we can increase total surplus by providing some consumer data to the seller. For example, for the consumers that gain from price personalization, providing their purchase history to the seller creates win-win interactions. Transferring consumer data to the seller can also generate a Pareto improvement when price personalization harms the consumer, if the associated gains in profits are such that the seller can compensate the consumer.

6.2 Price personalization with transfers

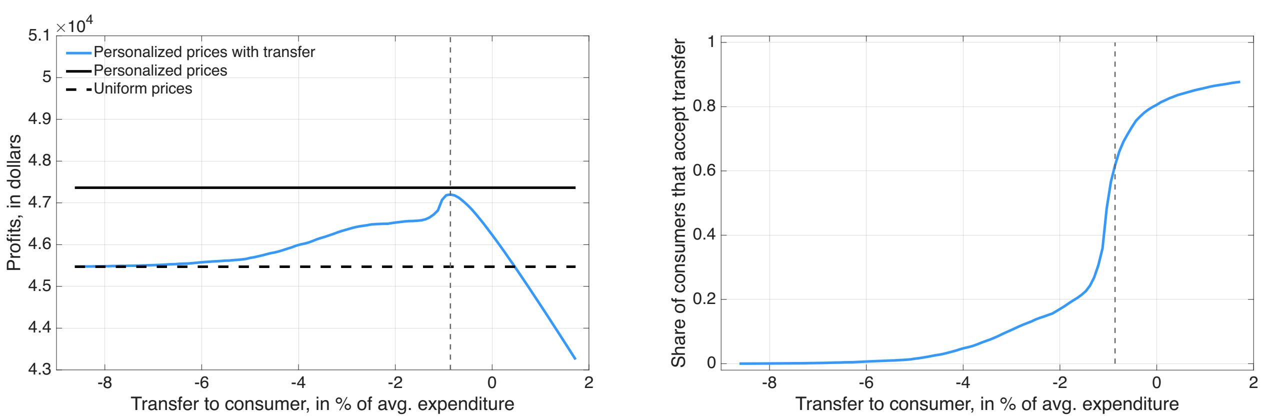
We consider an alternative pricing scheme where consumers have the option to provide or deny access to their personal data, for example by opting in or out of a loyalty program. The pricing game is in two stages. In the first stage, the seller offers a transfer to consumers in exchange of their consumer data. There is an asymmetry of information: the seller does not know the content of consumer histories before the transactions occur, hence it offers a unique price for data to all consumers. In the second stage, the seller sets personalized prices for consumers that shared their data and a uniform price to everyone else.

What is the optimal value of such transfer so as to maximize profits? On the one hand, purchasing consumer data could convince some price-insensitive consumers that are highly profitable to share their data. These consumers would then arbitrage the fixed payment against higher prices. On the other hand, asking for a membership fee could extract some surplus from price-sensitive consumers who may be willing to pay to access discounted prices. It would also deter price-sensitive consumers from sharing their data, since participation leads to higher prices on top of the membership fee. Which option is the most profitable is an empirical question we study in what follows.

We set up two separate experiments to understand what happens when the seller is allowed to set a two-part tariff as described above. In the first experiment, the seller uses the consumer data it acquired to personalize prices for these consumers, but it does not updates its belief about the remaining consumers that chose not to share their data. In this case, the resulting uniform prices are the same as before.

In the second experiment, we allow the seller to update its belief about the remaining consumers who refuse to provide their data. In this case, the seller creates a belief about the price sensitivity of the consumers that did not provide their data, and the uniform price is set conditionally on this belief. For the other consumers, they receive (or pay) the transfer and face personalized prices based on their purchase histories.

Figure 9: Price personalization with transfers (no belief updating from denying access)



(a) Profits with transfers (b) Share consumers that sell data

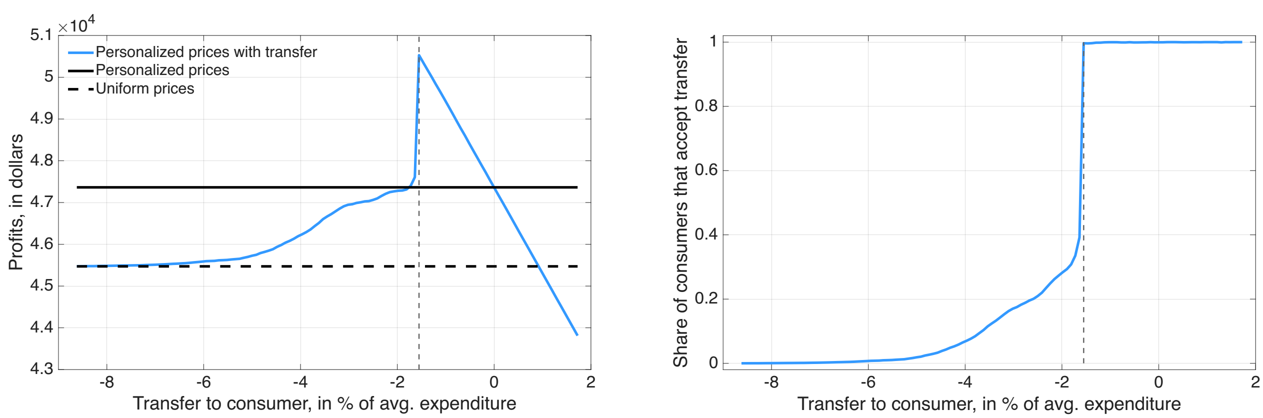
Notes: Panel (a) plots the profits under three different pricing regimes as a function of the amount of the transfer. Panel (b) plots the share of consumers that decide to sell their purchase history to the firm as a function of the transfer amount.

Experiment #1. The first set of results is available in Figure 9. We present the total seller profits (over all categories and all consumers) that arise from various transfer schemes. The transfer is a flat monthly fee paid by the seller (positive values) or the consumer (negative values), and we report its value as a percentage of the average consumer monthly expenditure ($11.60 per month).

The optimal transfer is a membership fee of 0.86% of the average consumer monthly basket, so consumers must pay to be part of the loyalty program. About 61% of the consumers accept to share their data in this context: they face a relatively low fee compared to the discounts they will receive from price personalization. Also, by construction, none of the consumers who pay supra-uniform prices under personalization share their data. The uniform prices therefore form an upper bound on prices, and all consumers that share their data receive infra-uniform prices.

The seller benefits from implementing this two-part tariff. First, it benefits from the market expansion effect from offering discounted prices to price-sensitive consumers. It is also able to extract more surplus from these consumers through the membership fee. As seen from panel (a) of Figure 9, the seller is almost able to reach the same level of profits using the two-part tariff versus the case where it has data on all consumers, covering 91% of the profit difference between the uniform case and the full personalization case.

Figure 10: Price personalization with transfers (with belief updating from denying access)



(a) Profits with transfers (b) Share consumers that accept transfer

Notes: Panel (a) plots the profits under three different pricing regimes as a function of the amount of the transfer. Panel (b) plots the share of consumers that decide to sell their purchase history to the firm as a function of the transfer amount.

Experiment #2. In the second experiment, we allow for the seller to updates its belief about consumers that refuse to share data. The results are presented in Figure 10. This change in the seller's behavior completely transform the results. We find that, for a given level of membership fee, more consumers share their data. Intuitively, this occurs because the seller understand that the consumers that do not share their data have low price sensitivity, and it increases the uniform prices accordingly.

We find that the market "unravels": as more consumers opt in, the uniform price increases, which convince more people to share data. In the resulting equilibrium, the seller charges a higher membership fee of 1.55% of the average consumer basket. Also, almost all consumers accept to share their data: the uniform price is the maximal price paid by the most price-insensitive group in each category and it forces consumers to accept the transfer since the fee is still relatively small compared to the discounted prices.

In this context, the firm can leverage the threat of high uniform prices to increase its profits beyond what is achievable in a perfect information context. Profits are 6.7% higher in this imperfect information scenario with two-part tariffs.

7 Conclusion

We propose a framework to study the value of consumer data in retail markets and show how firms can use Bayesian updating to personalize prices using only the information on consumers past purchases and a limited number of interactions with each consumer. This approach differs from studies focusing on pricing algorithms which typically require a large amount of data on consumers. In markets where consumer preferences evolve quickly, firms may want to focus only on the latest data available to form their pricing strategy.

We focus on a large supermarket in the U.S. and test the implementation of personalized pricing against the uniform pricing benchmark, in 24 product categories. We find that, even with few interactions with the consumers, the firm is able to reduce its uncertainty about consumer's price sensitivity by 30 to 60%. This leads to an overall increase in profits over all categories of around 4%. On the consumers' side, price personalization increases surplus slightly for a large number of price-sensitive consumers and harms severely a small number of price-insensitive consumers. The overall effect is mostly redistributive.

Finally, we use our framework to study a two-stage pricing game, where firms first propose a transfer to consumers to acquire their consumer data (e.g., opting in a loyalty program), and then personalize the prices for the consumers that opt in. In our context, the firm charges a membership fee and offers personalized discounts over the uniform price to consumers that opt in. Consumers then arbitrage the membership fee against the potential savings at the store from receiving personalized discounts.

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).*

Fox, J. T., il Kim, K., and Yang, C. (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics*, 195(2):236–254.

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.

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

Nevo, A. (2001b). 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 and Tables

Table A.1: Estimation of parameter support

Category	First-stage estimates				Support	
	a_1^k	a_2^k	b_1^k	b_2^k	supp(α^k)	supp(β^k)
Alcohol						
Beer	-0.73	0.78	-0.62	0.66	[-0.73, -0.10]	[-0.62, 0.04]
Wine	-0.74	1.03	-2.72	3.00	[-0.74, -0.10]	[-2.72, 0.28]
Dairy						
Butter/margarine/spreads	-1.48	1.09	-1.42	0.97	[-1.48, -0.39]	[-1.42, -0.45]
Cheese	-0.93	0.34	-1.30	1.06	[-0.93, -0.58]	[-1.30, -0.24]
Creams/creamers	-1.79	2.09	-1.39	1.69	[-1.79, -0.10]	[-1.39, 0.30]
Fresh eggs	-1.43	0.61	-1.57	0.74	[-1.43, -0.82]	[-1.57, -0.82]
Refrigerated juice/beverage	-0.90	0.60	-1.75	2.20	[-0.90, -0.30]	[-1.75, 0.45]
Milk	-1.74	0.45	-1.80	0.51	[-1.74, -1.29]	[-1.80, -1.29]
Yogurt	-1.59	1.35	-1.38	1.10	[-1.59, -0.24]	[-1.38, -0.28]
Frozen food						
Ice cream	-0.83	0.62	-0.84	0.61	[-0.83, -0.21]	[-0.84, -0.23]
Frozen pizza	-1.53	1.01	-1.47	2.78	[-1.53, -0.52]	[-1.47, 1.31]
Frozen potatoes/onions	-1.98	2.01	-1.02	0.85	[-1.98, -0.10]	[-1.02, -0.17]
Frozen vegetables	-1.49	0.49	-1.51	0.49	[-1.49, -1.00]	[-1.51, -1.02]
Grocery						
Cereal	-0.70	0.71	-2.30	2.30	[-0.70, -0.10]	[-2.30, 0.00]
Coffee	-1.02	0.95	-1.08	0.99	[-1.02, -0.10]	[-1.08, -0.09]
Crackers	-0.84	0.81	-0.89	0.84	[-0.84, -0.10]	[-0.89, -0.05]
Snacks grocery	-1.18	0.13	-1.81	0.80	[-1.18, -1.05]	[-1.81, -1.01]
Soft drinks/mixers	-1.48	0.25	-1.50	0.27	[-1.48, -1.23]	[-1.50, -1.23]
Water	-0.47	0.52	-0.59	0.63	[-0.47, -0.10]	[-0.59, 0.04]
Package meat						
Bacon	-1.48	0.54	-1.53	0.57	[-1.48, -0.94]	[-1.53, -0.96]
Dinner sausage	-1.22	0.46	-1.78	1.50	[-1.22, -0.76]	[-1.78, -0.28]
Taxable grocery						
Bleach/stain removers	-2.05	3.00	-1.05	1.94	[-2.05, -0.10]	[-1.05, 0.89]
Laundry detergent	-0.99	1.08	-0.81	2.66	[-0.99, -0.10]	[-0.81, 1.85]
Pet food	-1.33	1.24	-1.66	1.57	[-1.33, -0.10]	[-1.66, -0.08]

Notes: This table presents the results from the estimation of the support of α_i and β_i, as described in Section 4.3. The support is recovered from the estimates as $\text{supp}(\alpha^k) = [a_1^k, a_1^k + a_2^k]$ and $\text{supp}(\beta^k) = [b_1^k, b_1^k + b_2^k]$. To avoid computational problems related to positive or near zero price sensitivities, the upper bound of the support of α^k was replaced by -0.1 whenever the estimated upper bound was above -0.1.

Table A.2: Distribution of consumer types, by category

Dep: Alcohol
Cat: Beer

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.706	-0.292	0.162	0.494
2	-0.706	-0.292	0.560	0.197
3	-0.706	-0.292	0.936	0.115
4	-0.373	-0.572	0.162	0.098
5	-0.373	-0.572	0.560	0.043
6	-0.373	-0.572	0.936	0.038
7	-0.291	-0.422	0.162	0.015

Dep: Alcohol
Cat: Wine

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.573	0.195	0.162	0.535
2	-0.573	0.195	0.560	0.211
3	-0.573	0.195	0.936	0.127
4	-0.125	-0.502	0.162	0.072
5	-0.125	-0.502	0.560	0.029
6	-0.125	-0.502	0.936	0.027

Dep: Dairy
Cat: Butter/margarine/spreads

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.344	-1.395	0.162	0.415
2	-1.344	-1.395	0.560	0.129
3	-1.344	-1.395	0.936	0.024
4	-1.189	-0.478	0.936	0.058
5	-0.790	-1.368	0.162	0.191
6	-0.790	-1.368	0.560	0.111
7	-0.790	-1.368	0.936	0.038
8	-0.744	-0.501	0.936	0.034

Dep: Dairy
Cat: Cheese

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.910	-0.935	0.162	0.301
2	-0.910	-0.935	0.560	0.067
3	-0.910	-0.935	0.936	0.013
4	-0.819	-0.921	0.936	0.023
5	-0.744	-0.945	0.162	0.144
6	-0.744	-0.945	0.560	0.128
7	-0.721	-0.636	0.936	0.068
8	-0.671	-0.822	0.162	0.162
9	-0.641	-0.864	0.560	0.029
10	-0.629	-0.616	0.936	0.049
11	-0.612	-0.887	0.560	0.016

Dep: Dairy
Cat: Cream/creamers

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.750	-0.888	0.162	0.511
2	-1.750	-0.888	0.560	0.170
3	-1.750	-0.888	0.936	0.103
4	-1.230	-0.130	0.560	0.011
5	-1.230	-0.130	0.936	0.039
6	-1.158	-0.710	0.560	0.049
7	-1.139	0.148	0.162	0.095
8	-0.779	-1.378	0.560	0.010
9	-0.779	-1.378	0.936	0.011

Dep: Dairy
Cat: Fresh eggs

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.402	-1.309	0.936	0.035
2	-1.369	-1.450	0.162	0.447
3	-1.369	-1.450	0.560	0.125
4	-1.206	-0.930	0.560	0.027
5	-1.206	-0.930	0.936	0.024
6	-1.112	-0.939	0.560	0.034
7	-1.112	-0.939	0.936	0.063
8	-1.037	-1.485	0.162	0.106
9	-1.037	-1.485	0.560	0.041
10	-1.003	-0.902	0.936	0.012
11	-0.936	-1.339	0.162	0.054
12	-0.855	-1.112	0.560	0.013
13	-0.855	-1.112	0.936	0.021

Table A.3: Distribution of consumer types, by categories (cont.)

Dep: Dairy
Cat: Refrigerated juice/beverage

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.869	-0.987	0.162	0.482
2	-0.869	-0.987	0.560	0.141
3	-0.869	-0.987	0.936	0.049
4	-0.708	-0.999	0.560	0.028
5	-0.708	-0.999	0.936	0.017
6	-0.592	-0.769	0.162	0.064
7	-0.592	-0.769	0.560	0.056
8	-0.592	-0.769	0.936	0.059
9	-0.538	-0.375	0.162	0.036
10	-0.503	-0.331	0.560	0.015
11	-0.503	-0.331	0.936	0.029
12	-0.470	-0.950	0.162	0.024

Dep: Dairy
Cat: Milk

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.727	-1.730	0.162	0.350
2	-1.727	-1.730	0.560	0.074
3	-1.727	-1.730	0.936	0.016
4	-1.592	-1.773	0.162	0.050
5	-1.592	-1.773	0.560	0.024
6	-1.592	-1.773	0.936	0.020
7	-1.555	-1.537	0.560	0.051
8	-1.532	-1.429	0.936	0.042
9	-1.476	-1.760	0.162	0.206
10	-1.476	-1.760	0.560	0.075
11	-1.459	-1.305	0.936	0.039
12	-1.293	-1.708	0.560	0.016
13	-1.293	-1.708	0.936	0.038

Dep: Dairy
Cat: Yogurt

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.528	-1.002	0.162	0.494
2	-1.528	-1.002	0.560	0.184
3	-1.528	-1.002	0.936	0.096
4	-0.907	-1.195	0.162	0.112
5	-0.907	-1.195	0.560	0.056
6	-0.907	-1.195	0.936	0.058

Dep: Frozen food
Cat: Ice cream

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.816	-0.429	0.162	0.081
2	-0.762	-0.699	0.560	0.202
3	-0.723	-0.835	0.936	0.593
4	-0.626	-0.433	0.162	0.039
5	-0.532	-0.235	0.936	0.038
6	-0.532	-0.235	0.162	0.034
7	-0.439	-0.680	0.560	0.014

Dep: Frozen food
Cat: Frozen potatoes/onions

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.530	-1.322	0.162	0.513
2	-1.530	-1.322	0.560	0.179
3	-1.530	-1.322	0.936	0.093
4	-0.993	-1.340	0.162	0.075
5	-0.993	-1.340	0.560	0.061
6	-0.993	-1.340	0.936	0.051
7	-0.878	-1.164	0.162	0.018
8	-0.878	-1.164	0.936	0.010

Dep: Frozen food
Cat: Frozen pizza

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.943	-0.624	0.162	0.386
2	-1.943	-0.624	0.560	0.113
3	-1.943	-0.624	0.936	0.039
4	-0.669	-0.734	0.162	0.220
5	-0.669	-0.734	0.560	0.127
6	-0.669	-0.734	0.936	0.091
7	-0.453	-0.260	0.936	0.024

Table A.4: Distribution of consumer types, by category (cont.)

Dep: Frozen food
Cat: Frozen vegetables

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.427	-1.462	0.162	0.455
2	-1.427	-1.462	0.560	0.101
3	-1.427	-1.462	0.936	0.031
4	-1.260	-1.449	0.936	0.039
5	-1.223	-1.372	0.162	0.095
6	-1.223	-1.372	0.560	0.115
7	-1.195	-1.086	0.936	0.049
8	-1.093	-1.021	0.936	0.034
9	-1.080	-1.486	0.162	0.056
10	-1.080	-1.486	0.560	0.024

Dep: Grocery
Cat: Cereal

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.649	-0.680	0.162	0.472
2	-0.649	-0.680	0.560	0.155
3	-0.649	-0.680	0.936	0.052
4	-0.371	-0.216	0.936	0.054
5	-0.246	-0.715	0.162	0.121
6	-0.246	-0.715	0.560	0.085
7	-0.246	-0.715	0.936	0.028
8	-0.177	-0.051	0.936	0.020
9	-0.117	-0.478	0.162	0.013

Dep: Grocery
Cat: Coffee

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.965	-0.994	0.162	0.546
2	-0.965	-0.994	0.560	0.201
3	-0.965	-0.994	0.936	0.111
4	-0.556	-0.907	0.162	0.061
5	-0.556	-0.907	0.560	0.039
6	-0.556	-0.907	0.936	0.042

Dep: Grocery
Cat: Crackers

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.805	-0.644	0.162	0.542
2	-0.805	-0.644	0.560	0.186
3	-0.805	-0.644	0.936	0.067
4	-0.623	-0.056	0.560	0.013
5	-0.623	-0.056	0.936	0.068
6	-0.435	-0.865	0.162	0.053
7	-0.435	-0.865	0.560	0.041
8	-0.385	-0.097	0.936	0.018
9	-0.266	-0.798	0.162	0.011

Dep: Grocery
Cat: Snack grocery

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.172	-1.072	0.936	0.025
2	-1.167	-1.249	0.162	0.504
3	-1.167	-1.249	0.560	0.155
4	-1.167	-1.249	0.936	0.015
5	-1.113	-1.016	0.936	0.020
6	-1.062	-1.034	0.560	0.074
7	-1.052	-1.230	0.162	0.102
8	-1.046	-1.052	0.560	0.011
9	-1.046	-1.052	0.936	0.094

Dep: Grocery
Cat: Soft drinks/mixers

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.460	-1.470	0.162	0.313
2	-1.460	-1.470	0.560	0.115
3	-1.460	-1.470	0.936	0.045
4	-1.373	-1.307	0.162	0.045
5	-1.357	-1.433	0.162	0.173
6	-1.352	-1.239	0.162	0.030
7	-1.352	-1.239	0.560	0.108
8	-1.352	-1.239	0.936	0.084
9	-1.277	-1.502	0.936	0.012
10	-1.244	-1.371	0.162	0.045
11	-1.244	-1.371	0.560	0.016
12	-1.244	-1.371	0.936	0.012

Table A.5: Distribution of consumer types, by category (cont.)

Dep: Grocery
Cat: Water

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.451	-0.455	0.162	0.546
2	-0.451	-0.455	0.560	0.180
3	-0.451	-0.455	0.936	0.060
4	-0.310	-0.070	0.936	0.033
5	-0.237	0.002	0.560	0.059
6	-0.237	0.002	0.936	0.042
7	-0.165	-0.480	0.162	0.060
8	-0.129	-0.149	0.936	0.019

Dep: Packaged meat
Cat: Bacon

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.454	-1.389	0.162	0.497
2	-1.454	-1.389	0.560	0.161
3	-1.454	-1.389	0.936	0.060
4	-1.208	-1.428	0.162	0.094
5	-1.208	-1.428	0.560	0.079
6	-1.208	-1.428	0.936	0.076
7	-1.093	-1.051	0.936	0.018
8	-1.083	-1.367	0.162	0.015

Dep: Packaged meat
Cat: Dinner sausage

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.200	-1.184	0.162	0.434
2	-1.200	-1.184	0.560	0.130
3	-1.200	-1.184	0.936	0.043
4	-1.062	-1.253	0.936	0.017
5	-0.998	-0.834	0.560	0.010
6	-0.998	-0.834	0.936	0.044
7	-0.978	-1.186	0.162	0.150
8	-0.978	-1.186	0.560	0.089
9	-0.945	-0.940	0.936	0.029
0	-0.862	-0.829	0.936	0.021
11	-0.839	-1.268	0.162	0.022
12	-0.839	-1.268	0.560	0.011

Dep: Taxable grocery
Cat: Bleach/stain remover

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.932	-0.825	0.162	0.502
2	-1.932	-0.825	0.560	0.164
3	-1.932	-0.825	0.936	0.070
4	-0.591	-0.741	0.560	0.033
5	-0.591	-0.741	0.936	0.054
6	-0.383	-0.872	0.162	0.104
7	-0.383	-0.872	0.560	0.043
8	-0.383	-0.872	0.936	0.017
9	-0.314	0.412	0.936	0.012

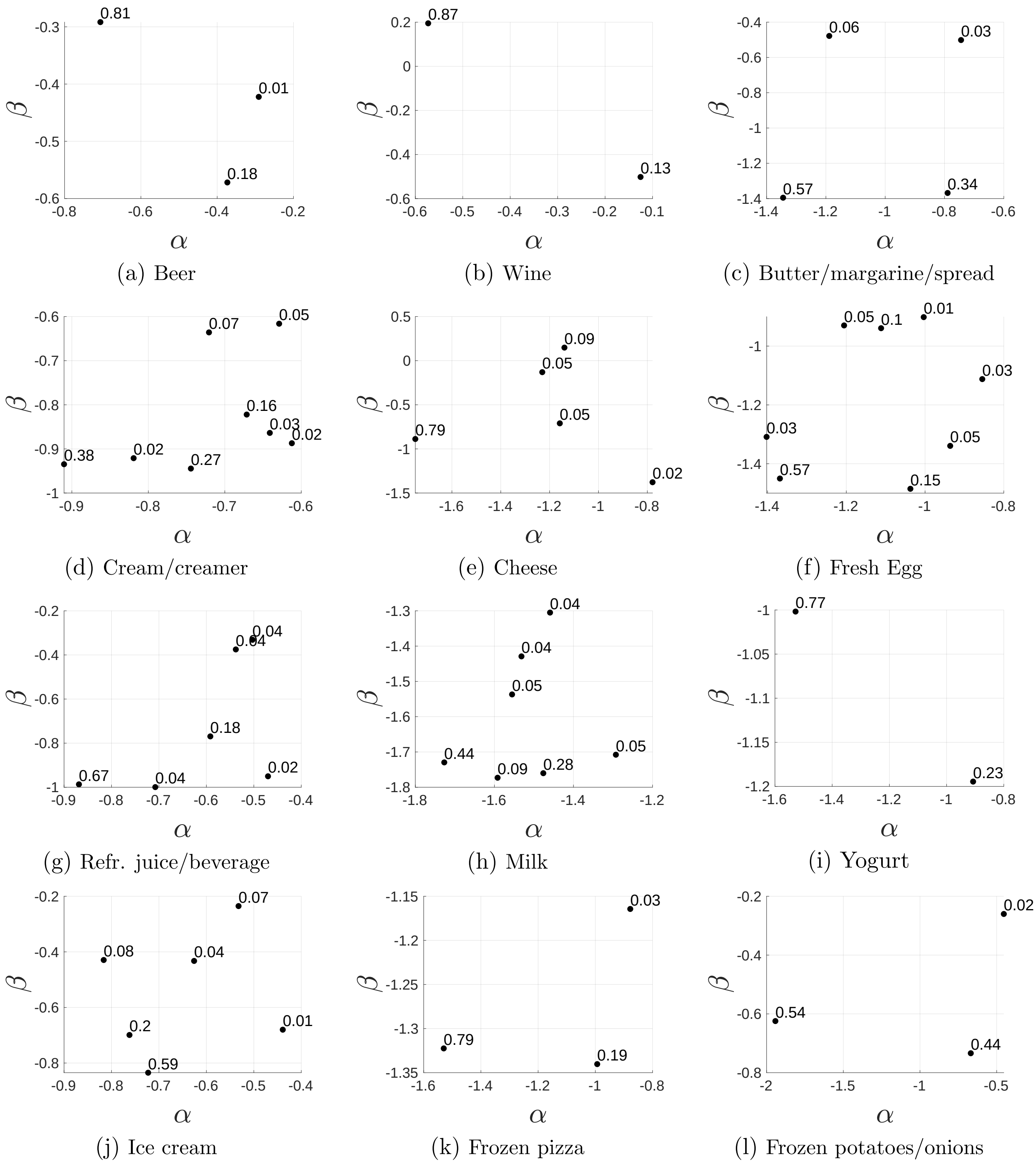
Dep: Taxable grocery
Cat: Laundry detergent

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-0.992	-0.561	0.162	0.548
2	-0.992	-0.561	0.560	0.202
3	-0.992	-0.561	0.936	0.107
4	-0.448	-0.615	0.560	0.038
5	-0.448	-0.615	0.936	0.047
6	-0.415	-0.717	0.162	0.058

Dep: Taxable grocery
Cat: Pet food

Type ID	α_d	β_d	ρ_d	ϕ_d
1	-1.260	-1.476	0.162	0.553
2	-1.260	-1.476	0.560	0.209
3	-1.260	-1.476	0.936	0.122
4	-0.560	-1.319	0.162	0.054
5	-0.560	-1.319	0.560	0.031
6	-0.560	-1.319	0.936	0.031

Figure A.1: Marginal distributions of consumer preferences



(a) Beer (b) Wine (c) Butter/margarine/spread

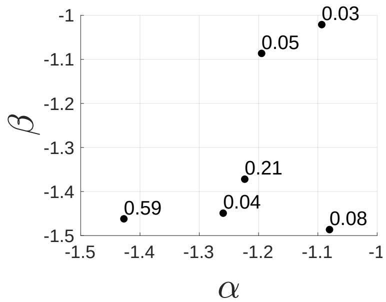
(d) Cream/creamer (e) Cheese (f) Fresh Egg

(g) Refr. juice/beverage (h) Milk (i) Yogurt

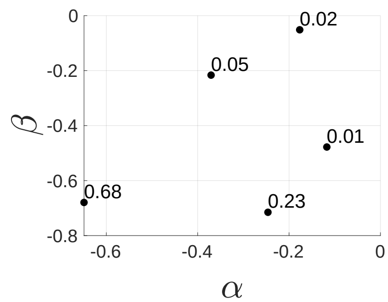
(j) Ice cream (k) Frozen pizza (l) Frozen potatoes/onions

Notes This figure represents the marginal distribution of consumers preferences, integrated over the probability to get a purchase occasion. The values on the graphs represent the probability mass distribution.

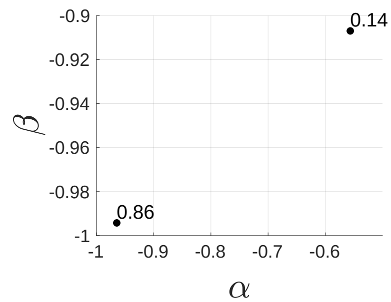
Figure A.2: Marginal distributions of consumer preferences (cont.)



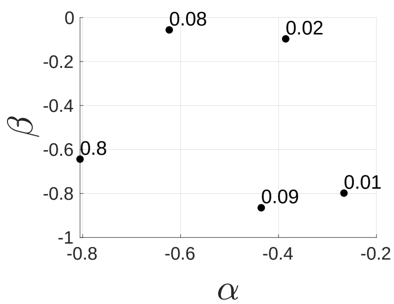
(a) Frozen vegetables



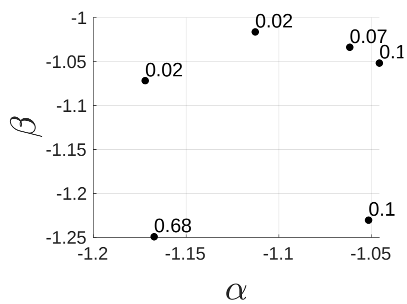
(b) Cereal



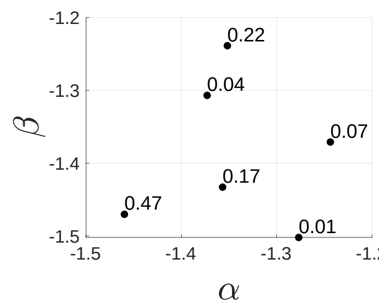
(c) Coffee



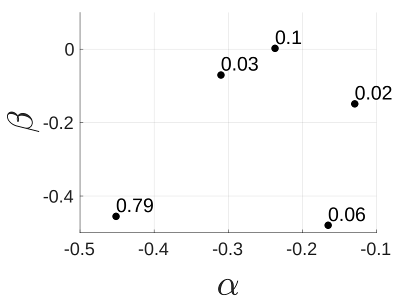
(d) Cracker



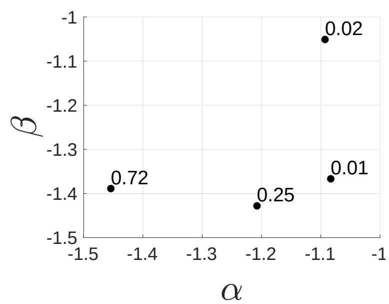
(e) Snack grocery



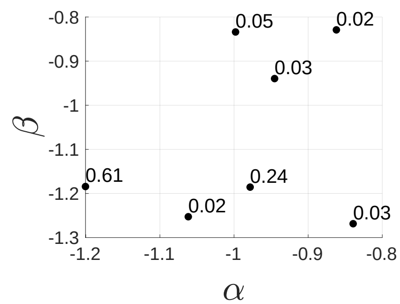
(f) Soft drink/mixer



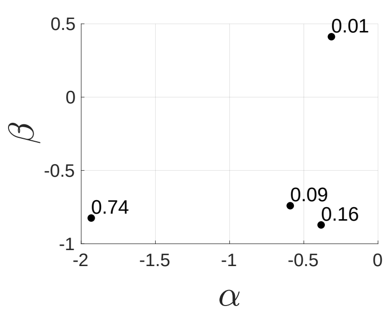
(g) Water



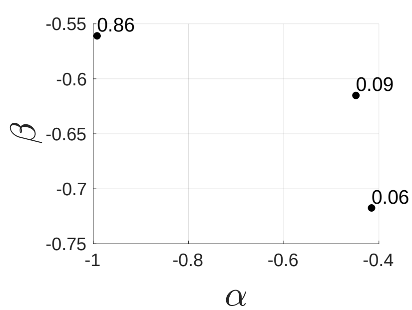
(h) Bacon



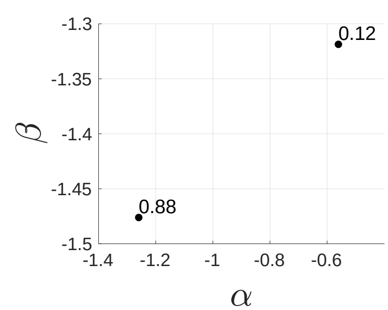
(i) Dinner sausage



(j) Bleach/stain Remover



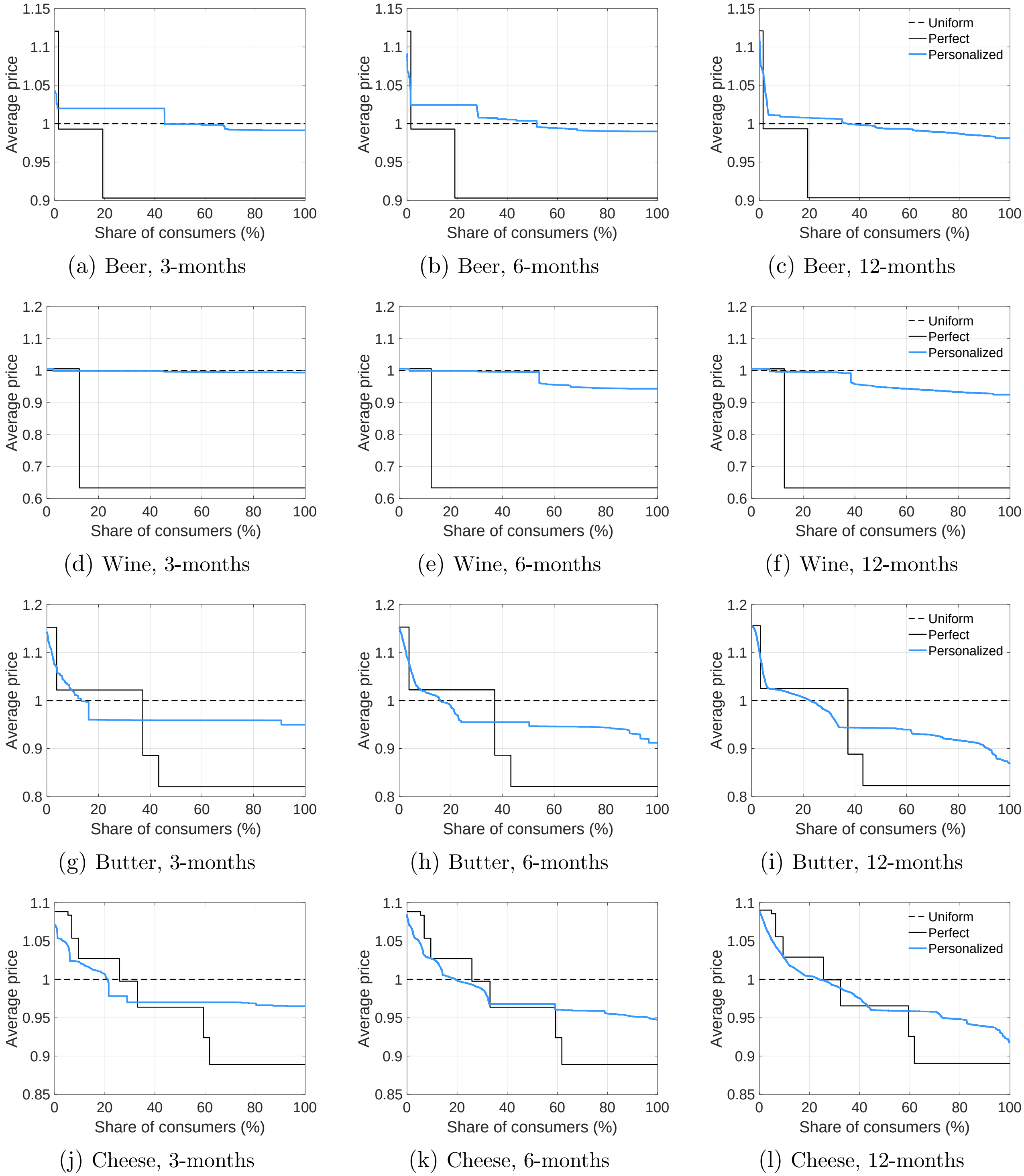
(k) Laundry detergent



(l) Pet food

Notes This figure represents the marginal distribution of consumers preferences, integrated over the probability to get a purchase occasion. The values on the graphs represent the probability mass distribution.

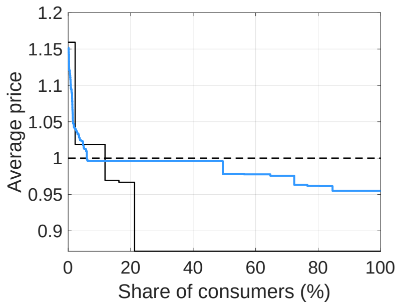
Figure A.3: Price dispersion, by history length



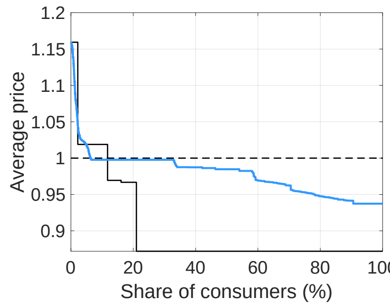
(a) Beer, 3-months
(b) Beer, 6-months
(c) Beer, 12-months
(d) Wine, 3-months
(e) Wine, 6-months
(f) Wine, 12-months
(g) Butter, 3-months
(h) Butter, 6-months
(i) Butter, 12-months
(j) Cheese, 3-months
(k) Cheese, 6-months
(l) Cheese, 12-months

Notes: Distributions of prices over the distribution of consumers, after observing a purchase history of three, six, and twelve months respectively.

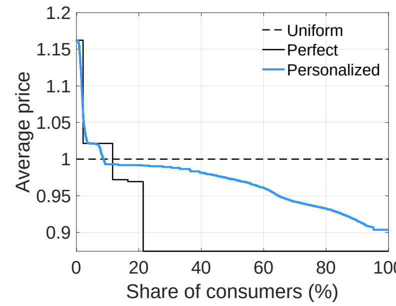
Figure A.4: Price dispersion, by history length (cont.)



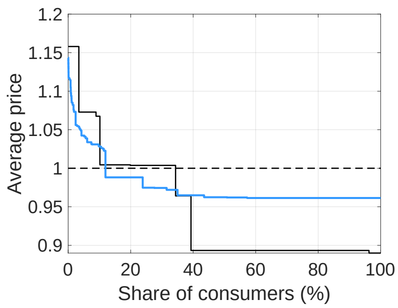
(a) Cream, 3-months



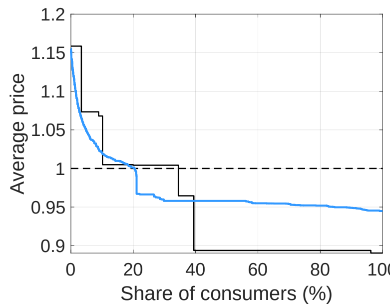
(b) Cream, 6-months



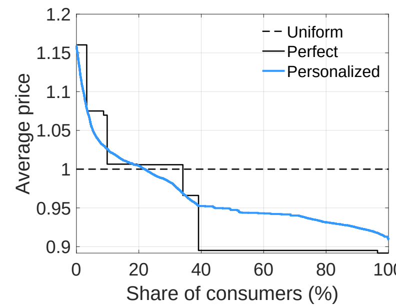
(c) Cream, 12-months



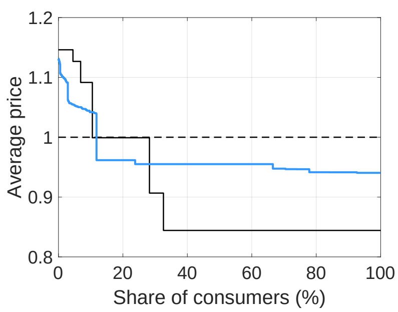
(d) Fresh eggs, 3-months



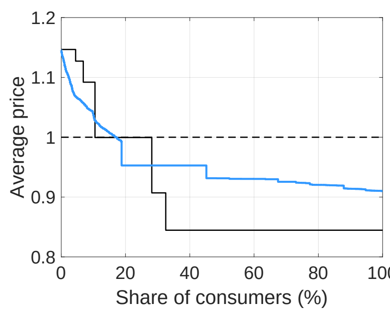
(e) Fresh eggs, 6-months



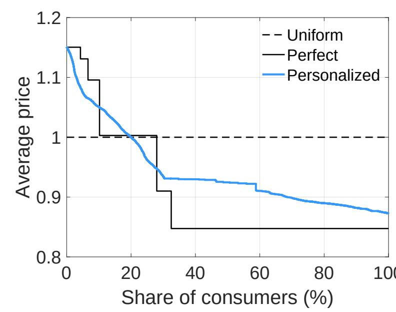
(f) Fresh eggs, 12-months



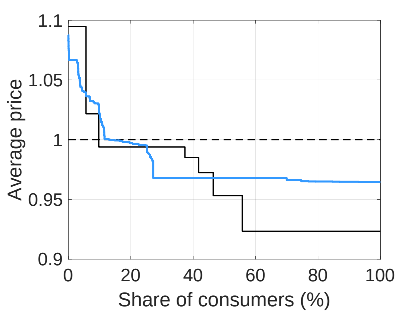
(g) Refr. Juice, 3-months



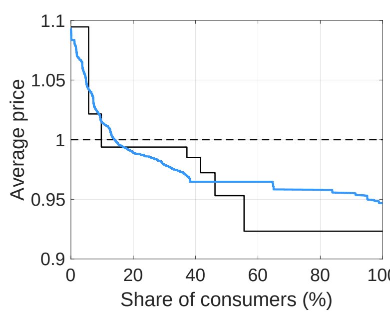
(h) Refr. Juice, 6-months



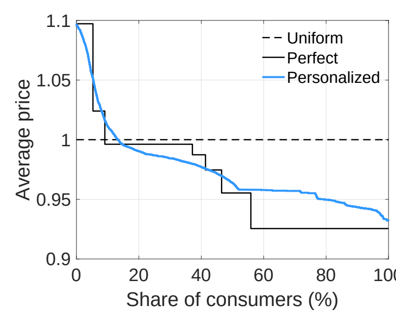
(i) Refr. Juice, 12-months



(j) Milk, 3-months



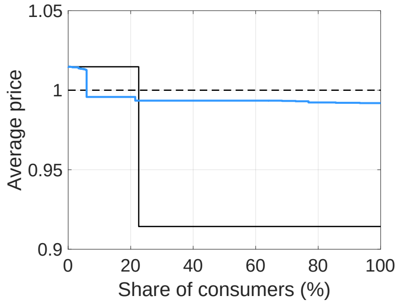
(k) Milk, 6-months



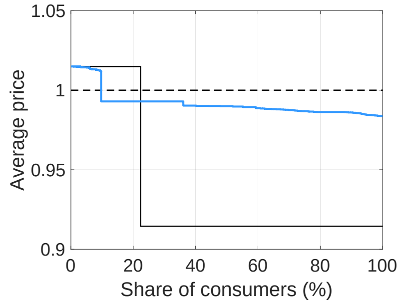
(l) Milk, 12-months

Notes: Distributions of prices over the distribution of consumers, after observing a purchase history of three, six, and twelve months respectively.

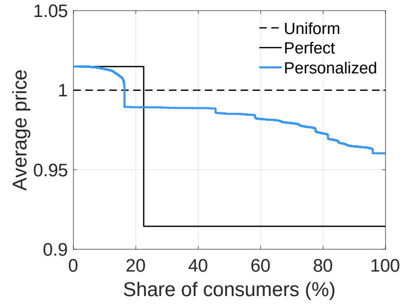
Figure A.5: Price dispersion, by history length (cont.)



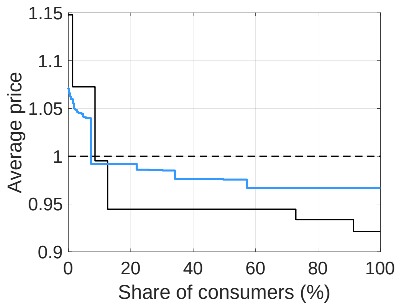
(a) Yogurt, 3-months



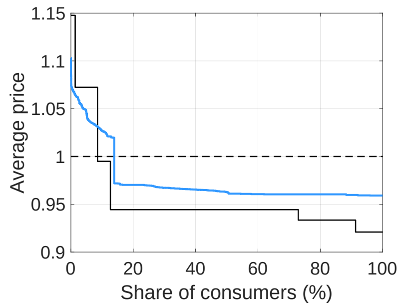
(b) Yogurt, 6-months



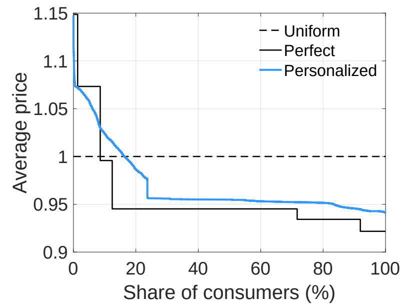
(c) Yogurt, 12-months



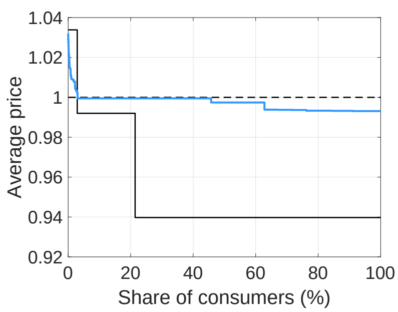
(d) Ice cream, 3-months



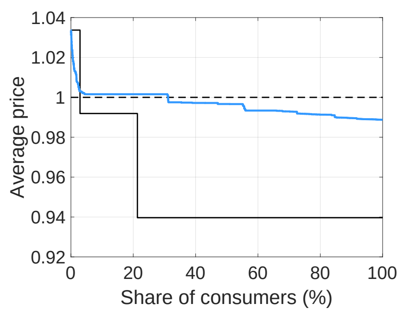
(e) Ice cream, 6-months



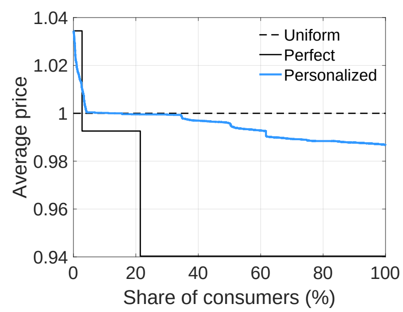
(f) Ice cream, 12-months



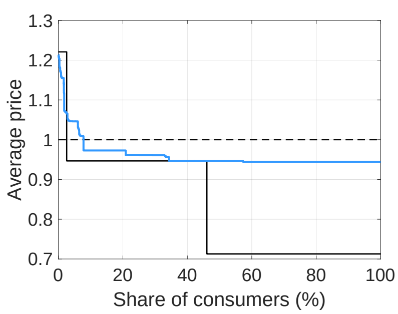
(g) Fz. Pizza, 3-months



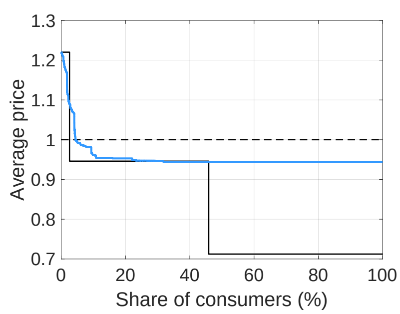
(h) Fz. Pizza, 6-months



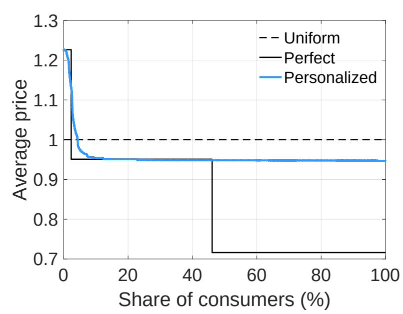
(i) Fz. Pizza, 12-months



(j) Fz. Potatoes, 3-months



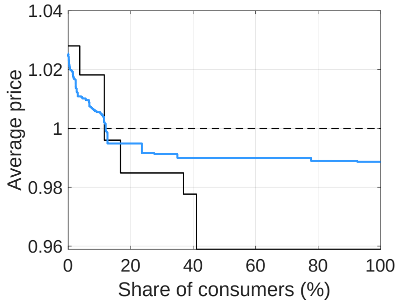
(k) Fz. Potatoes, 6-months



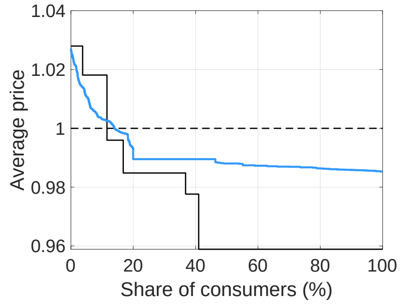
(l) Fz. Potatoes, 12-months

Notes: Distributions of prices over the distribution of consumers, after observing a purchase history of three, six, and twelve months respectively.

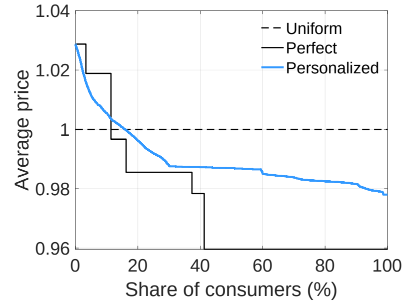
Figure A.6: Price dispersion, by history length (cont.)



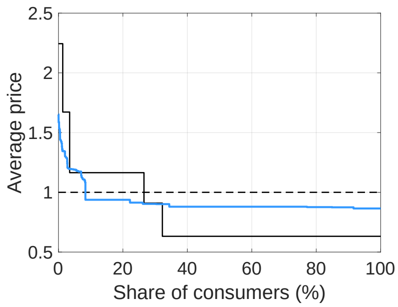
(a) Fz. Vegetables, 3-months



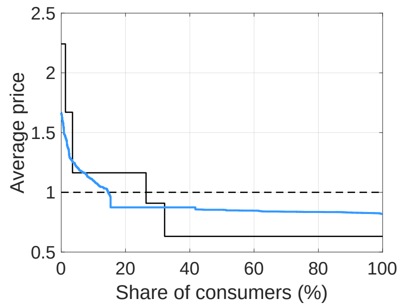
(b) Fz. Vegetables, 6-months



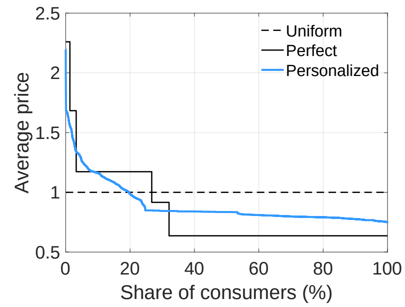
(c) Fz. Vegetables, 12-months



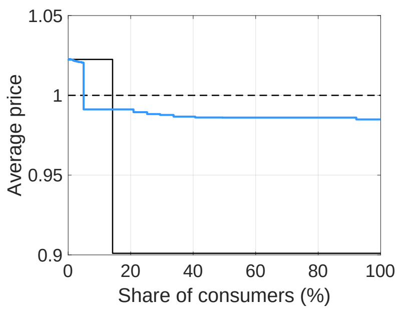
(d) Cereal, 3-months



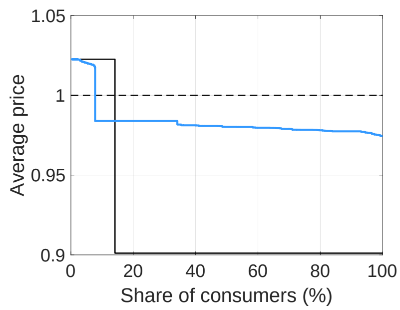
(e) Cereal, 6-months



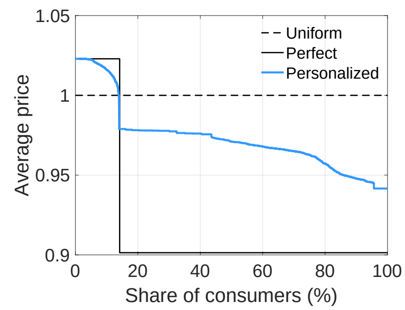
(f) Cereal, 12-months



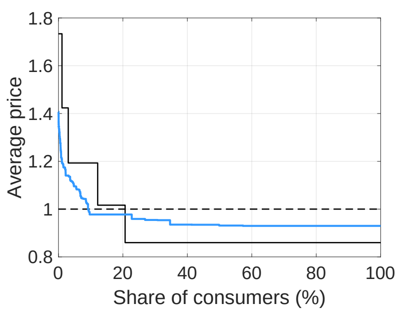
(g) Coffee, 3-months



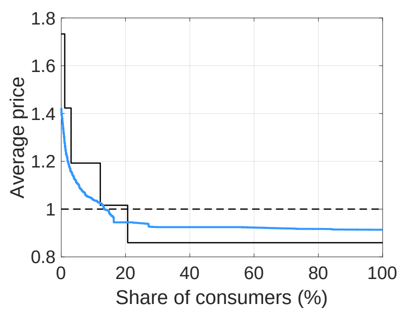
(h) Coffee, 6-months



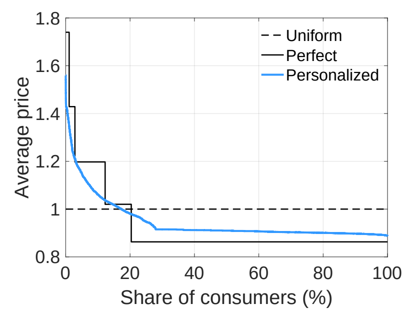
(i) Coffee, 12-months



(j) Crackers, 3-months



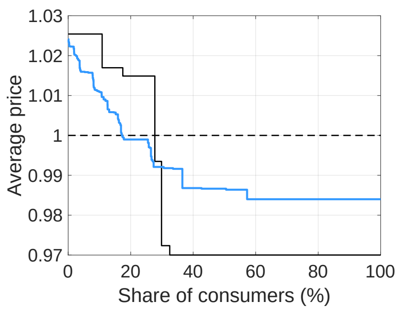
(k) Crackers, 6-months



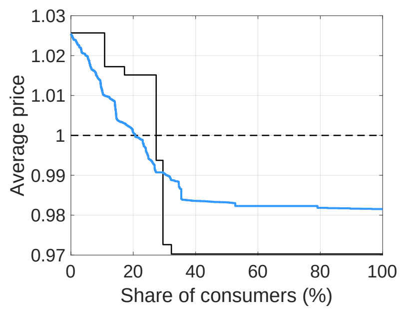
(l) Crackers, 12-months

Notes: Distributions of prices over the distribution of consumers, after observing a purchase history of three, six, and twelve months respectively.

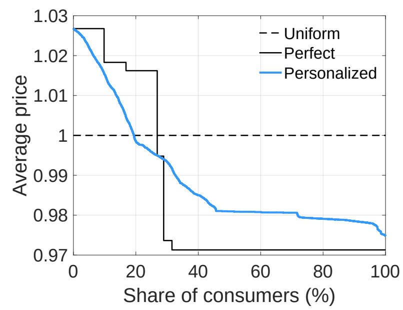
Figure A.7: Price dispersion, by history length (cont.)



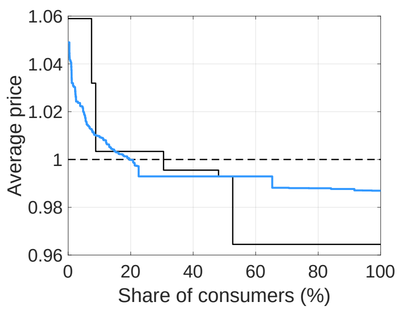
(a) Snacks, 3-months



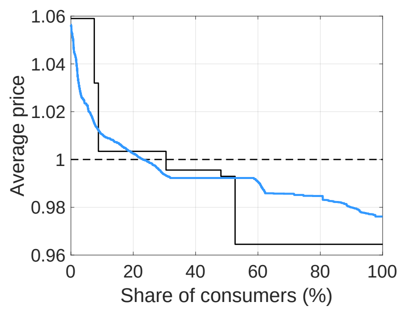
(b) Snacks, 6-months



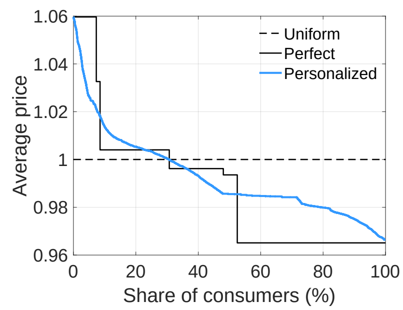
(c) Snacks, 12-months



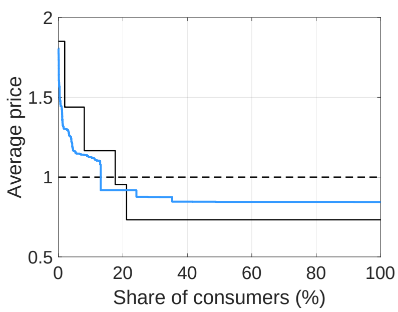
(d) Soft drinks, 3-months



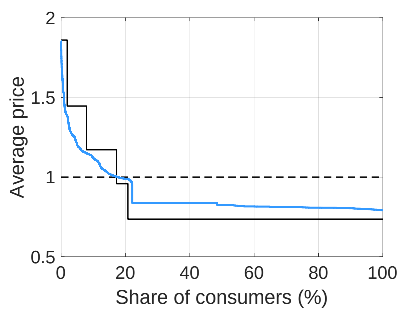
(e) Soft drinks, 6-months



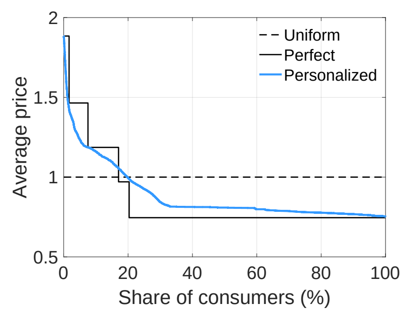
(f) Soft drinks, 12-months



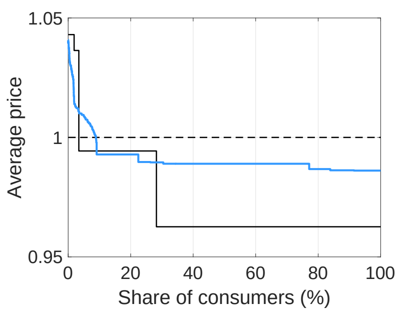
(g) Water, 3-months



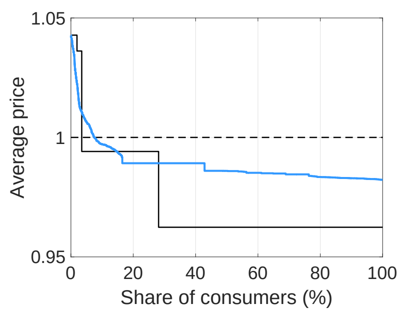
(h) Water, 6-months



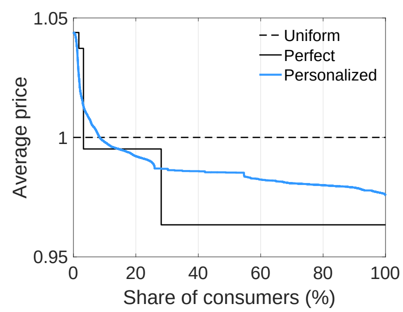
(i) Water, 12-months



(j) Bacon, 3-months



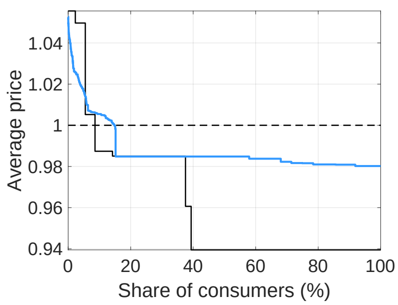
(k) Bacon, 6-months



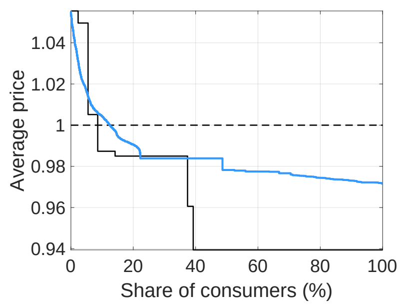
(l) Bacon, 12-months

Notes: Distributions of prices over the distribution of consumers, after observing a purchase history of three, six, and twelve months respectively.

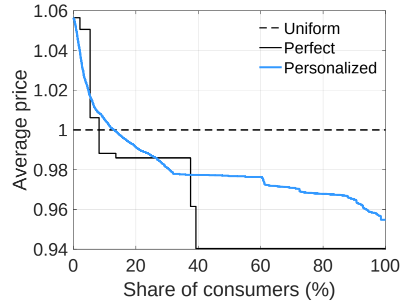
Figure A.8: Price dispersion, by history length (cont.)



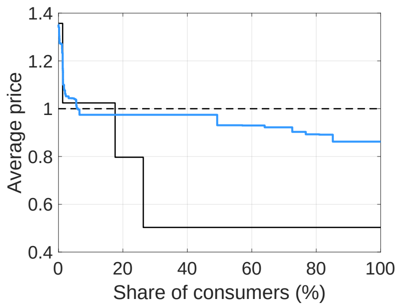
(a) Dinner sausage, 3-months



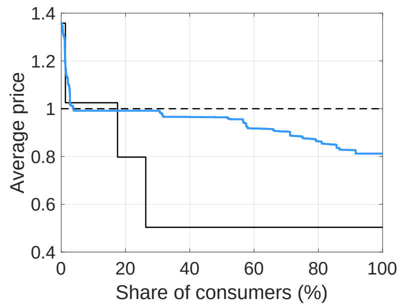
(b) Dinner sausage, 6-months



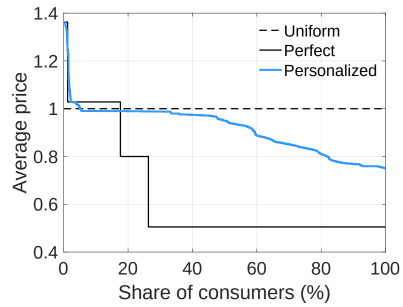
(c) Dinner sausage, 12-months



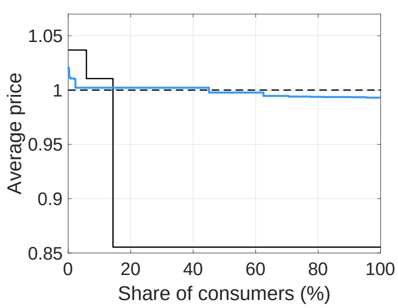
(d) Bleach, 3-months



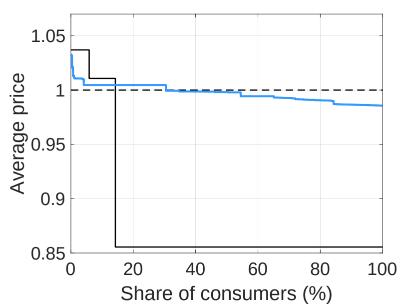
(e) Bleach, 6-months



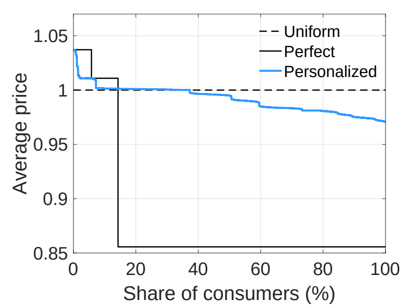
(f) Bleach, 12-months



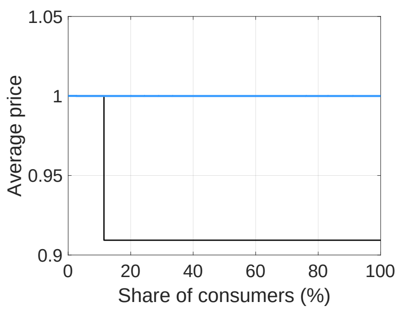
(g) Laundry, 3-months



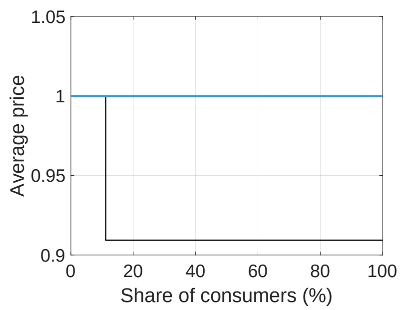
(h) Laundry, 6-months



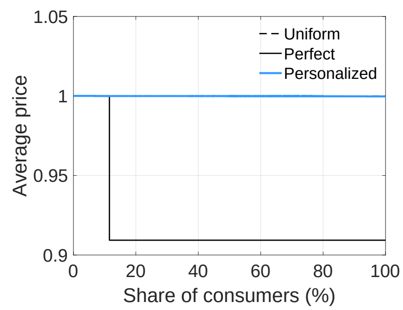
(i) Laundry, 12-months



(j) Pet food, 3-months



(k) Pet food, 6-months



(l) Pet food, 12-months

Notes: Distributions of prices over the distribution of consumers, after observing a purchase history of three, six, and twelve months respectively.

B Details on the Data

B.1 Data

The data used in our empirical application come from DecaData.[22] 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, for each transaction, at the universal product code level (henceforth UPC).

ODYSSEY database. The ODYSSEY database describes all purchases realized at the 853 stores. The data covers all sales that occur between January 1 and December 31, 2018. The data includes the date of transaction, a store identifier, the UPC of the products purchased, the quantity purchased (in units or weight), the transaction price, as well as a transaction identifier that allows us to identify joint purchases. Additionally, we observe a consumer identifier for a subset of consumers who are enrolled in a loyalty card program. For these consumers, it is possible to track their purchases over time, but not across stores, as the identifiers are specific to consumer-store pairs.

ARGO database. The ARGO database contains wholesale purchases from the 853 retailers and include the date of delivery, the store identifier, the UPC, the quantity ordered and delivered (these could differ if the product is back-ordered), as well as the unit wholesale price of the goods delivered. The database spans December 2017 and all of 2018.

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). The PRODUCT database does not include observed product characteristics (e.g., the nutritional content, the amount of sugar, calories) which limits our modeling decisions. Products characteristics have been widely used to estimate preferences and to construct instruments to identify the price coefficient in demand estimation since Berry et al. (1995).

[22] Source: `https://decadata.io/data`.

B.2 Sample selection

We focus on one representative supermarket to perform the analysis. We chose a store in the vicinity of the median store in terms of total revenues. Our chosen store had \$13.1 million dollars in sales in 2018, placing it in the 52$^{\text{nd}}$ percentile of the distribution. We restrict our attention to the 17,756 consumers enrolled in a loyalty program and for whom we can track purchases over time.[23] These consumers account for approximately 80% 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 quality of the match between the retail sales data and the wholesale purchases data, and we drop departments for which we were unable to recover the wholesale price for more than 20% of products. Wholesale prices are not explicitly used in the analysis. We find that they are still useful for assessing the fit of our model by comparing observed and predicted wholesale prices.

We restrict our attention to the largest food categories within the chosen food departments, based on revenues. Categories that account for less than 5% of total revenues by food department are removed.[24] This selection rule leaves us with 24 food categories.[25] Finally, we remove a small number of UPC in each category for which we were unable to normalize units of measurement into standardized quantities. These account for approximately 2.6% of the remaining sample in terms of total revenues.

Table B.1 provides some statistics on consumers' purchases as we restrict the sample. We first look at the average consumer spending per transaction. Comparing column (1) and (2) suggests that consumers who enrolled in a loyalty program do not differ much from other consumers in terms of total purchases per transaction: the average spending is 30.4 versus 30.8 dollars per transaction and the quartiles of the distribution are also close to each other. Consumers within our sample spend on average 46.5 dollars per month on groceries at our

[23] We eliminated a small number of customer with unreasonably high monthly purchases. We assume these to be institutional or private customers, such as restaurants or other types of non-household customers.

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

[25] Initially, our analysis included baby formula as an additional category. Our estimation routine performed poorly on this category, usually returning a unique consumer type, which made our study of price discrimination trivial in this case. For this reason, we removed this category from the analysis.

Table B.1: Per consumer spending for representative store

	Full dataset	Selected consumers	Selected departments	Selected categories	Final sample
Consumer spending per transaction					
Mean	30.8	30.4	20.8	12.2	11.9
S.D.	39.4	34.3	24.1	12.1	11.8
Min.	1.0	1.0	0.0	0.0	0.0
25th pct.	7.1	8.0	5.2	3.8	3.8
Median	16.4	18.0	12.0	8.0	8.0
75th pct.	37.6	39.0	26.3	16.0	15.8
Max.	230.0	182.2	129.7	63.0	61.3
Observations	397,377	335,130	295,339	217,045	215,419
Consumer spending per month					
Mean	–	46.5	27.9	11.9	11.6
S.D.	–	88.5	56.4	24.6	24.0
Min.	–	0.0	0.0	0.0	0.0
25th pct.	–	0.0	0.0	0.0	0.0
Median	–	0.0	0.0	0.0	0.0
75th pct.	–	54.2	29.3	12.3	12.0
Max.	–	471.9	308.3	133.2	130.2
Observations	–	213,072	213,072	213,072	213,072
Consumer spending per month, conditional on a visit					
Mean	–	101.0	63.5	30.5	29.8
S.D.	–	107.4	70.6	31.3	30.6
Min.	–	0.2	0.3	0.4	0.4
25th pct.	–	25.6	14.6	8.2	8.0
Median	–	62.1	37.1	19.5	19.0
75th pct.	–	136.8	85.3	41.5	40.5
Max.	–	471.9	308.3	133.2	130.2
Observations	–	97,953	93,565	83,162	82,845
Total sales (million)	13.07	10.41	6.27	2.68	2.61
# of consumers	All	17,756	17,756	17,756	17,756
Store ID	1697				

Notes: All reported values are in 2018 USD, except when noted otherwise. We report basic statistics on consumer spending per transaction, consumer spending per month, and consumer spending per month conditional on the consumer visiting the store in a given month. Column (1) reports statistics computed using the full dataset. Column (2) reports statistics computed using consumers enrolled in a loyalty program only. Column (3) further restrict the sample to our chosen food departments. Column (4) further restrict the sample to our chosen categories. Column (5) presents statistics for the final sample used in the empirical application.

chosen store. The median spending per month is 0, meaning that less that half of consumers visit the store each month. Conditionally on a visit, consumers spend on average 101 dollars on groceries, and the median consumer spends 62.1 dollars per month.

When considering our most restricted dataset which includes 6 food departments and 24 categories, consumers in our sample spend on average 29.8 dollars per month, conditionally on a visit, and the median consumer spends 19 dollars per month. Our final sample captures approximately one quarter of all purchases made by registered consumers, or one quarter of consumers' monthly purchases.

B.3 Product definition and aggregation

We consider that each product category (e.g., cereal) forms a separate market and the various manufacturers (e.g., Kellogg, General Mills, etc.) constitute separate, differentiated products within each market. To limit the proliferation of products with very low market shares, we group all manufacturers with a revenue share below 5% into a single fringe product.

Table B.2 provides a breakdown of the various product categories and the associated number of products (including the fringe product). We compare the market share in terms of revenues and quantity sold of the top selling non-fringe product, the bottom selling non-fringe product, and the fringe product in each category. Except for the case of the wine category which features a large number of very small producers, we find that the fringe products are not the market leader and have a market share comparable to other products in their respective categories.

The typical category features between 2 and 7 differentiated products, with an average of 4.5. Since we aggregate over UPC in different package sizes, we normalize the quantities purchased, retail prices, and wholesale prices in terms of the most common package size in each product category. We then aggregate all UPC that fall under the same product definition, and the retail and wholesale prices for that product are the weighted average of the underlying UPC-specific prices. A summary is available in Table 1. We encountered a few occurrences of UPC with negative sales (e.g., refunds) or negative margins (e.g., price below wholesale price). We discard these observations, which account for approximately 0.2% of the sample (before aggregation).

Quantities and retail prices are available from the main ODYSSEY database. Wholesale

Table B.2: Product definition and size of fringe product

Category	# Products	% of revenues			% of quantity sold		
		Max.	Min.	Fringe	Max.	Min.	Fringe
Alcohol							
Beer	4	47.7	7.1	20.6	45.5	7.6	24.5
Wine	7	20.6	5.8	35.6	23.9	5.9	27.8
Dairy							
Butter/margarine/spreads	5	26.0	21.2	5.4	37.5	15.1	4.2
Cheese	4	56.7	7.9	6.7	50.3	9.7	11.8
Creams/creamers	4	38.8	25.5	1.8	34.0	32.0	1.9
Fresh eggs	2	76.6	23.4	0	80.0	20.0	0
Refrigerated juice/beverage	4	41.8	20.5	15.6	36.9	20.0	17.7
Milk	5	71.4	5.6	10.9	66.1	4.0	10.8
Yogurt	4	48.9	8.2	12.1	42.2	5.2	16.2
Frozen food							
Ice cream	4	55.3	11.8	3.7	60.5	12.6	2.7
Frozen pizza	4	39.9	13.3	0	46.0	14.7	0
Frozen potatoes/onions	4	49.5	15.1	0.6	51.1	16.4	0.7
Frozen vegetables	4	62.5	8.9	6.0	53.1	11.1	4.2
Grocery							
Cereal	5	38.3	7.7	4.1	39.0	12.7	8.3
Coffee	6	40.2	8.8	10.4	46.3	4.2	8.6
Crackers	5	42.1	11.3	1.3	32.7	11.3	2.1
Snacks grocery	3	67.1	11.3	21.6	44.1	25.3	30.6
Soft drinks/mixers	5	35.0	12.4	0.7	35.0	20.1	1.5
Water	6	40.6	5.3	8.8	43.1	2.9	6.7
Packaged meat							
Bacon	6	24.1	6.1	13.1	24.2	4.2	16.9
Dinner sausage	6	45.3	5.5	21.1	46.4	4.6	25.3
Taxable grocery							
Bleach/stain removers	4	71.9	5.8	2.7	78.7	2.9	2.2
Laundry detergent	4	66.2	8.5	6.6	54.9	7.2	10.0
Pet food	4	51.2	10.1	3.6	49.9	9.0	5.5

Notes: This table presents a breakdown of the various product categories and the associated number of products. Column (4) and (5) display the share of total revenues for the top and bottom selling brand (excl. the fringe product) for each category. Column (6) displays the share of total revenues for the fringe product for each category. Column (7) and (8) display the share of total quantity sold for the top and bottom selling products (excl. the fringe product) for each category. Column (9) displays the share of total quantity sold for the fringe product for each category. All values are in percentages.

prices are constructed from wholesale purchases, detailed in the ARGO database. We cannot precisely link retail sales with wholesale purchases in both databases. We computed the average wholesale price for each UPC in the ARGO database in each month and imputed that value to all retail sales of that UPC observed in the same month. We believe that this is an adequate proxy for the wholesale price: we capture either the cost of the most recent retail purchases of that product if the retail sale occurs near the end of the month or the cost of restocking the same product whenever the retail sale occurs at the beginning of the month. When no wholesale purchases are observed in the same month a product is sold (this can happen for non-perishables), we use the wholesale price computed from the previous month instead.

C Details on the Estimation

C.1 Estimation of the support

We begin by showing that the model presented in equation (12) is a special case of equation (11). We omit the category index k to simplify the notation. The likelihood of the sample is

$$\begin{aligned}\mathscr{L}(\cdot) &= \sum_{i=1}^{M} \log\left(\Pr(\mathbf{h}_i)\right), \\ &= \sum_{i=1}^{M} \log\left(\sum_{d=1}^{D} \phi_d \prod_{t=1}^{T}\left[\prod_{j=0}^{J} \mathbb{P}(\theta_d, \boldsymbol{\delta}_t)^{Y_{ijt}}\right] \rho_d^{\sum_{j=0}^{J} Y_{ijt}} \left(1-\rho_d\right)^{1-\sum_{j=0}^{J} Y_{ijt}}\right).\end{aligned} \tag{14}$$

For the estimation of the support, we assume that:

A1. Each consumer has a unique type (e.g., $d = i$) with equal weight, hence $\phi_i = 1/M$; and

A2. $\rho_i = \rho$, $\forall i = 1, ..., M$, ρ is independent of $\theta_i = (\alpha_i, \beta_i)$.

Under these assumptions, equation (14) becomes

$$\begin{aligned}\mathscr{L}(\cdot) &= \sum_{i=1}^{M} \log\left(\prod_{t=1}^{T}\left[\prod_{j=0}^{J} \mathbb{P}(\theta_i, \boldsymbol{\delta}_t)^{Y_{ijt}}\right] \rho^{\sum_{j=0}^{J} Y_{ijt}} \left(1-\rho\right)^{1-\sum_{j=0}^{J} Y_{ijt}}\right), \\ &= \sum_{i=1}^{M} \log\left(\prod_{t=1}^{T}\prod_{j=0}^{J} \mathbb{P}(\theta_d, \boldsymbol{\delta}_t)^{Y_{ijt}}\right) + \kappa\end{aligned}$$

where

$$\kappa = \sum_{i=1}^{M} \log\left(\prod_{t=1}^{T} \rho^{\sum_{j=0}^{J} Y_{ijt}}\right) + \sum_{i=1}^{M} \log\left(\prod_{t=1}^{T} (1-\rho)^{1-\sum_{j=0}^{J} Y_{ijt}}\right).$$

Notice that κ is constant by assumption, and we can ignore its contribution to the conditional likelihood.

Similarly for the constraint, we have that

$$\begin{aligned} s_{jt} &= \sum_{d=1}^{D} \phi_d \frac{\rho_d \cdot \exp(\delta_{jt} + \beta_d + \alpha_d p_{jt})}{1 + \sum_{j'} \exp(\delta_{j't} + \beta_d + \alpha_d p_{j't})} \\ &= \rho \cdot \frac{1}{M} \sum_{i=1}^{M} \frac{\exp(\delta_{jt} + \beta_i + \alpha_i p_{jt})}{1 + \sum_{j'} \exp(\delta_{j't} + \beta_i + \alpha_i p_{j't})} \end{aligned} \tag{15}$$

The constraints on ϕ_d hold trivially by assumption **A1**.

The estimation is based around finding an appropriate support for α_i and β_i. We impose the following parametric assumptions,

$$\alpha_i = a_1 + a_2\, \eta_{i1},$$

$$\beta_i = b_1 + b_2\, \eta_{i2},$$

where (η_{i1}, η_{i2}) are distributed as independent Uniform(0, 1), following the suggestion in Fox et al. (2016). The support of each parameter can then be recovered from the estimated coefficients $(\boldsymbol{a}, \boldsymbol{b})$ as

$$\text{supp}(\alpha_i) = [a_1, a_1 + a_2],$$

$$\text{supp}(\beta_i) = [b_1, b_1 + b_2].$$

To avoid complications related to positive or near zero price sensitivities, we replace the upper-bound of the support of α_i by -0.1 whenever the upper-bound is above -0.1. The results of the estimation of the support of the preference parameters are available in Table A.1.

C.2 Estimation on the fixed grid

We estimate the model separately for each category k. In what follows, we omit the category index k for simplicity. The estimated parameters are the type probabilities, $\{\phi_d\}_{d=1,\ldots,D}$, and the mean utilities of products $\{\boldsymbol{\delta}_t\}_{t=1,\ldots,T}$, with $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{JT})$. The preference parameters (α_d, β_d) and the probability to obtain a shopping occasion ρ_d are provided on a fixed grid.

The grid is constructed by interacting 20 preference types, drawn from the joint support of α_i and β_i using Halton draws (see Table A.1) with the three purchase occasion types estimated from the data on purchase occasions (see Table 2). Therefore, the initial grid is comprised of 60 different consumer types in each category.

Notice that nothing in the estimation routine restricts consumers preference patterns across categories. For example, some consumers may exhibit a high price sensitivity in one category and a low price sensitivity in another. However, since ρ_d is tied to the sequence of purchase occasions each consumer experiences, and that this sequence does not vary across categories, each consumer should have the same ρ_d across categories (this will create some correlation between consumer types across categories).

We estimate the following optimization problem,

$$\max_{\{\phi_d\}_{d=1,\dots,D},\,\{\boldsymbol{\delta}_t\}_{t=1,\dots,T}} \sum_{i=1}^{M} \log\left(\sum_{d=1}^{D}\phi_d \prod_{t=1}^{T}\left[\prod_{j=0}^{J}\mathbb{P}_{jt}(\theta_d,\boldsymbol{\delta}_t)^{Y_{ijt}}\right]\rho_d^{\sum_{j=0}^{J}Y_{ijt}}(1-\rho_d)^{1-\sum_{j=0}^{J}Y_{ijt}}\right),$$

$$\text{s.t.}\quad s_{jt} = \sum_{d=1}^{D}\phi_d\,\frac{\rho_d\cdot\exp(\delta_{jt}+\beta_d+\alpha_d p_{jt})}{1+\sum_{j'}\exp(\delta_{j't}+\beta_d+\alpha_d p_{j't})},$$

$$\Pr(\text{Occasional shopper}) = \sum_{d=1}^{D}\phi_d\,\mathbb{1}(d\in\text{Occasional shopper}),$$

$$\Pr(\text{Regular shopper}) = \sum_{d=1}^{D}\phi_d\,\mathbb{1}(d\in\text{Regular shopper}),$$

$$\Pr(\text{Frequent shopper}) = \sum_{d=1}^{D}\phi_d\,\mathbb{1}(d\in\text{Frequent shopper}),$$

$$\sum_{d=1}^{D}\phi_d = 1,\quad \phi_d\geq 0,$$

where the individual contribution to the likelihood is the probability to observe each consumer's purchase history over 12 months, and we restrict the marginal distribution of ρ_d to match the unconditional distribution estimated by finite mixture, see Table 2.

After estimating the model once, several types return near zero probability mass (i.e., $\phi_d \approx 0$). We remove iteratively these types from the grid and redo the estimation until all types have a probability mass above 0.01. Formally, we first remove types from the 20 grid points over (α_i, β_i) for which all three combinations with ρ_i returned probability masses summing to less than 0.01. We perform the estimation again and repeat this step until we obtain a stable grid in the (α_i, β_i)-space. We interact that grid with the three point of support for ρ_i and remove types with a probability mass less than 0.01. We perform the estimation again and repeat this step until we obtain a stable grid in the $(\alpha_i, \beta_i, \rho_i)$-space. The resulting grid is the final grid, and the associated type probabilities $\{\phi_d\}_{d=1,...,D}$ are the chosen estimates.

We want to point out that this iterative removal procedure does not change our results in any meaningful way. We do this to avoid the proliferation of types and reduce the number of irrelevant computation we are required to do in counterfactual experiments (e.g., computing personalized prices for consumer types with zero mass).