

October, 2026

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

Isis Durrmeyer, Jean-François Fournel and Mario Samano

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

Isis Durrmeyer[†] Jean-François Fournel[‡] Mario Samano[§]

October 11, 2026

Abstract

Grocery loyalty programs record purchase histories that retailers do not use to personalize prices. We measure how much they forgo with a structural model of demand and pricing in which a Bayesian seller learns preferences from past purchases and personalizes prices. Profits rise by 7.1% with one-year histories, capturing two thirds of the gains under perfect knowledge of preferences. Most consumers (84.1%) gain, but the losses of the minority outweigh these gains, and aggregate consumer surplus falls. An optimal loyalty program charges a membership fee, yet most consumers enroll, because opting out signals price insensitivity.

Keywords: personalized pricing, behavior-based price discrimination, purchase histories, Bayesian learning, loyalty programs, value of information

JEL Classifications: D42, D83, L11, L81, D12.

*We acknowledge financial support from the European Research Council under grant ERC-2019-STG-852815 “PRIDISP” and Agence Nationale de Recherche under grant ANR-17-EUR-0010 (Investissements d’Avenir program). We thank Hester Zhang, Felix Schleef, Carlos Santos and participants of the Toulouse School of Economics 15th Digital Economics Conference, the 59th Canadian Economics Association Annual Conference, Paris IO Day 2026, the 2nd Markets and Waves Conference, and the 53rd EARIE Conference for their comments.

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

[‡]Harvard University, Salata Institute for Climate and Sustainability. E-mail: jfournel@fas.harvard.edu

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

1 Introduction

Personalized pricing enables firms to charge consumer-specific prices that reflect heterogeneous preferences, moving beyond uniform pricing toward first-degree price discrimination. Personalized pricing requires two conditions: knowing each consumer’s preferences, and being able to charge each consumer a different price. Digital technologies have made both possible. Loyalty programs and online accounts record each consumer’s purchases over time, which should in principle reveal preferences. Conditioning consumers’ prices on their purchase histories is possible through a loyalty card or app in a brick-and-mortar store and through the account itself in an online store.¹ Competition and consumer protection authorities have examined the consequences of such pricing for consumers and are currently shaping enforcement policies addressing personalized pricing.²

Grocery retail is where these data are richest. Nearly 70% of US consumers use loyalty cards at grocery stores, groceries are bought often enough for histories to record many choices per consumer, and major chains such as Kroger and Safeway already use personalized promotions and discounts through mobile apps.³ Yet personalized pricing has not materialized at scale in this industry. We verify this in our data, which cover all purchases at a US grocery store: for a given product on a given day, virtually all consumers pay the same price, and the rare exceptions are quantity discounts and discounts for cardholders.

How much money are supermarkets leaving on the table? More generally, as digital channels make personalized pricing feasible, who gains and who loses from price discrimination based on purchase histories? We address these questions by treating purchase histories as an economic asset, and personalized pricing as the strategy that monetizes it. We consider a counterfactual in which the seller uses the purchases its loyalty program already records to update its beliefs about each consumer’s preferences, and sets consumer-specific prices based on those beliefs.

¹Throughout the paper, we use “personalized pricing” to refer specifically to pricing strategies that condition on purchase histories, that is, behavior-based price discrimination.

²See the OECD (2018) roundtable “Personalised Pricing in the Digital Era”, <https://www.oecd.org/daf/competition/personalised-pricing-in-the-digital-era.htm>, and the European Parliament (2022) study “Personalised Pricing”, [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/734008/IPOL_STU\(2022\)734008_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/734008/IPOL_STU(2022)734008_EN.pdf). As recently as August 2026, the FTC sought public comment on a set of new policies, see “Federal Trade Commission’s Proposed Enforcement Policy Statement Regarding Personalized Pricing”, <https://www.ftc.gov/news-events/news/press-releases/2026/08/ftc-seeks-comment-enforcement-policy-statement-regarding-personalized-pricing>.

³Chain Store Age, “Grocery shoppers continue relying on loyalty programs, apps, discounts”, <https://chainstoreage.com/grocery-shoppers-continue-relying-loyalty-programs-apps-discounts>.

The value of purchase histories depends on (i) what the seller learns about preferences from histories of a given length, (ii) how valuable is the information once prices are personalized, and (iii) the terms on which consumers agree to share their data.⁴

We develop a structural empirical model of a monopolistic multi-product seller to represent a supermarket’s pricing decisions. We specify a discrete choice model of grocery purchases across 24 product categories, each containing differentiated products varying in quality and cost. We allow for consumer heterogeneity along three dimensions: shopping occasion probability, category valuation, and price sensitivity. Together, the three parameters define a consumer’s type. We do not impose a parametric distribution of types: within each category, types are drawn from a discrete distribution whose support and weights we estimate. The shopping occasion probability is common to all categories, while the category valuation and the price sensitivity are category-specific. We place no restriction on how they relate across categories, so a consumer may be price-sensitive in one category but not in another.

Under personalized pricing, the supermarket updates its beliefs about each consumer’s type by Bayes’s rule, combining the population distribution of types with the consumer’s observed purchases, and sets the consumer’s prices on the posterior. Learning operates category by category, so the supermarket updates its beliefs about a consumer’s type in a category from that consumer’s history in that category alone. For two pairs of frequently co-purchased categories, we show that pooling their histories would add little to the seller’s gains.

We estimate the model on proprietary transaction data from a US supermarket. The data record every purchase made at the store in 2018 by the 20,748 cardholders of its loyalty program, so we observe one year of purchase histories. The store’s shipment records also give us the wholesale price of every product. We approximate the distribution of types by a discrete distribution, as in Fox et al. (2011) and Fox et al. (2016), and estimate it by maximum likelihood. The unit of observation is a consumer’s purchase history over the year. Histories link a consumer’s choices across shopping occasions and reveal persistent heterogeneity in preferences. Market share constraints identify the mean utilities of products in each month, as

⁴Policymakers note that the evidence on these questions is limited. In August 2026 the FTC wrote “The limited economic research on the question suggests that while personalized pricing is likely to increase business profits, benefits to some consumers are accompanied by losses to other consumers and that the more sophisticated personalized pricing practices become, the less likely consumers are to benefit.”, see https://www.ftc.gov/system/files/ftc_gov/pdf/p034101-ftc-enforcement-policy-statement-re-personalized-pricing-proposed-for-public-comment.pdf.

in [Dubé et al. \(2012\)](#) and [Grieco et al. \(2025\)](#). The likelihood identifies the dispersion of price sensitivities but not their level. We recover the level from the supply side. The supermarket sets optimal uniform prices given wholesale prices, and we choose the level so that the margins implied by the model match observed margins on average. The recovered marginal costs of individual products, which are not fitted, track the observed wholesale prices closely.

We use the estimated model to trace out the value of purchase histories as they accumulate, varying history length from 1 to 36 months. Purchase histories are informative but learning is slow and incomplete. One year of data reduces the seller’s uncertainty about consumers’ price sensitivity by 19% on average across categories and about their category taste by 39%. Three years of data reduce them by 34% and 55%, respectively. The seller’s profits, however, accrue much faster.

Relative to uniform pricing, profits increase by 3.8% with three months of data, by 7.1% with one year, and by 9.2% with three years. What pays is identifying the few price-insensitive consumers to surcharge, and purchase histories deliver that information within months. These gains are not far below the maximum attainable: even with perfect knowledge of every consumer’s preferences, price discrimination would raise the seller’s profits by 10.8%. One-year histories already capture 65.9% of the potential gain and three-year histories capture 85.2%. These magnitudes help explain why supermarkets do not personalize prices. With 7.1% of profits at stake after a year of data, and 10.8% under full information, modest frictions such as implementation costs, consumer backlash, or regulatory risk can rationalize the uniform pricing strategy.

These gains assume that the seller knows the distribution of consumer types and observes which product each consumer buys. Most of the gains remain when either assumption is relaxed. A seller that updates from a misspecified prior retains 73.0% to 89.4% of the gain with one-year histories. A seller that observes only whether a purchase is made each month retains 80%.

In our model, the seller learns from histories generated at uniform prices, a by-product of its sales, and consumers do not manipulate their purchases to obtain lower personalized prices later. The seller could instead distort prices to make histories more informative, at the cost of current profits. We solve for the price experiment that maximizes the seller’s profits after learning, and for the prices of a seller that maximizes the sum of current and next-period profits. The experimentation motive turns out to be small. The best experiment raises the

seller's subsequent profits by at most 1.5% of the uniform-pricing profits. The profits forgone during the experiment would take more than two years to recover even in the most favorable category. A seller that maximizes the sum of current and next-period profits keeps every price within 0.5% of the uniform price. Consumers, for their part, gain little from withholding all their purchases in a category to obtain lower prices: in 23 of the 24 categories, no consumer recovers the forgone surplus within a year of lower prices.

Personalized pricing raises total welfare but lowers consumer surplus. Consumers lose about 39 cents for every dollar of profit the seller gains, and aggregate consumer surplus falls by 0.7% of total consumer expenditure. This loss is carried by very few consumers. After one-year histories, between 85% and 98% of consumers, depending on the category, pay less than the uniform price, while the few identified as price-insensitive face premia that can exceed 100% of the uniform price. Surcharges transfer surplus from price-insensitive consumers to the seller, while discounts bring in price-sensitive consumers who would not buy at uniform prices. As a result, 84.1% of consumers are better off, while the losses of the price-insensitive minority outweigh the many small gains.

Finally, purchase histories must come from consumers themselves, who might refuse to surrender data that can be used to price against them. We model the decision to enroll in a loyalty program and solve for the transfer that maximizes the seller's profit. Enrollees face personalized prices and pay or receive a monthly transfer, while non-enrollees face a uniform price. We consider two scenarios. In the first, non-enrollees face the baseline uniform price. The seller charges a membership fee, 25.8% of consumers enroll, and profits rise by 4.2% relative to uniform pricing. In the second scenario, the seller treats opting out as a signal of price insensitivity and raises the prices of non-enrollees. Most consumers then enroll to avoid these prices, profits rise by 11.0% and exceed the personalized pricing benchmark. The program leaves 66% of consumers worse off than under uniform pricing. Voluntary enrollment therefore protects consumers only when the seller does not learn from their refusal.

Relation to the literature. The theory of behavior-based price discrimination considers pricing when sellers recognize their past customers. It builds on the analysis by [Hart and Tirole \(1988\)](#) of a monopolist who learns a buyer's valuation from past purchases, and was developed for competitive markets by [Villas-Boas \(1999\)](#) and [Fudenberg and Tirole \(2000\)](#).⁵

⁵See [Fudenberg and Villas-Boas \(2006\)](#) for a survey.

This literature characterizes the market conditions under which such price discrimination is profitable (Villas-Boas, 2004; Acquisti and Varian, 2005). More generally, Bergemann et al. (2015) characterize the divisions of surplus between the seller and consumers that price discrimination can induce. This theoretical literature is however silent on the size of the profit gain, the data it requires, and which consumers gain or lose, and we supply these magnitudes for one retail market. Our loyalty-program results connect to the theoretical literature on the economics of privacy, which characterizes when consumers agree to share their data if they anticipate price discrimination (for example, Conitzer et al., 2012, Ichihashi, 2020 and Ali et al., 2023).⁶ These models take the gains and losses from personalization as given, whereas we estimate them and derive enrollment and the optimal fee from the estimates.

We contribute to the empirical literature on personalized pricing, which is thin because the practice is rarely observed and its effects are quantified through model-based counterfactuals. Some existing studies measure the profit gains from price personalization with different consumer data: web-browsing data at Netflix (Shiller, 2020), firms’ characteristics and their responses to a price experiment on a job-matching platform (Dubé and Misra, 2023), characteristics of graduate students for college tuition (Waldfogel, 2015), repeated purchases of pickup trucks (Sagl, 2023), and lenders’ credit scores for personalized credit limits and interest rates (Matcham, 2025). Kehoe et al. (2020) study experience goods, where consumers learn their own tastes as they consume. These studies take the data as given, in length and in content, whereas we make the data the object of study.

Closest to us are two studies on pricing conditional on purchase histories in grocery retail. In Rossi et al. (1996), a manufacturer offers households personalized coupons on one brand of tuna, tailored to their posterior preferences obtained from purchase histories, in the same spirit as our study. Smith et al. (2023) follow the same design for one brand of mayonnaise and compare the pricing policies derived from Bayesian hierarchical choice models with those derived from machine learning demand models. The hierarchical models prove at least as profitable and far more reliable. Neither paper solves the seller’s pricing problem: retail prices are taken from the data, and the personalized discount is the only decision. Two further differences are specific to these studies. In both, the instrument is a discount, so no consumer loses. In our setting prices themselves are personalized, and some consumers pay more. In addition, we work on the scale of a store rather than a category.

⁶See Acquisti et al. (2016) for a survey.

Our results also contribute to the literature on the consequences of uniformity of grocery prices (DellaVigna and Gentzkow, 2019, Hitsch et al., 2021). These studies document nearly uniform prices within chains and dispersion across chains. Miravete (2026) finds that store-level pricing in place of uniform chain pricing can reduce welfare. Adams and Williams (2019) find that uniform pricing across stores can raise oligopolists’ profits, because a common price commits them to compete less aggressively. Miravete et al. (2024) explain the absence of discrimination across local markets by implementation costs that must exceed the forgone profits. In the same way, our profit gains bound the costs of discriminating across consumers on their purchase histories. Customers who see the price of an item they bought fall afterwards cut subsequent purchases (Anderson and Simester, 2010). Consumers who learn that others pay less may react in the same way, and fear of backlash leads firms to conceal personalization inside seemingly impersonal pricing strategies, such as dynamic pricing (Shiller, 2025).

Finally, we contribute to the literature on demand estimation across multiple product categories. Eizenberg et al. (2021) and Thomassen et al. (2017) focus on the choice of a supermarket and simplify the choice of products within it. Like Döpper et al. (2025), Atalay et al. (2026), and Bhattacharya et al. (2026), we abstract from supermarket choice to focus on within-category differentiation across many categories. We differ in how we model heterogeneity. We exploit loyalty program data, which track each consumer’s choices over time, to estimate a discrete mixture that leaves the joint distribution of preferences unrestricted. Persistent heterogeneity is what the seller learns from purchase histories, so restricting its distribution would predetermine what histories can reveal. The mixture also protects our welfare conclusions from functional form dependence: a single logit generates welfare gains from price discrimination mechanically (Miravete, 2026), and a mixture with an unrestricted type distribution does not.

The remainder of the paper is organized as follows. Section 2 presents the model of demand, seller learning, and pricing. Section 3 presents the data and documents price uniformity. Section 4 describes the estimation method, the identification strategy, and the estimation results. Section 5 quantifies the speed of learning, the dispersion of personalized prices, their profitability, and the consequences for consumer surplus and welfare. It also measures the value of price experiments for the seller and consumers’ incentive to withhold purchases during the learning phase. Section 6 studies distributional effects and the optimal design of a loyalty program where consumers who opt in share their purchase data to obtain personalized prices.

Section 7 concludes.

2 The Model

We describe a model of demand and supply in which heterogeneous consumers make repeated purchases from a monopolistic, multi-product seller.⁷ The model has two phases. In the observation phase, the seller knows only the distribution of consumers’ preferences in the population, not individual-specific preferences; it sets optimal uniform prices and observes consumers’ purchases. Consumers are myopic: they choose the product that maximizes their current utility, without distorting their purchases to influence the seller’s beliefs.⁸ In the personalization phase, the seller applies Bayes’s rule to update its beliefs about each consumer’s preferences and sets personalized prices based on the updated beliefs.

2.1 Demand

Setup. Demand arises from heterogeneous consumers indexed by $i = 1, \dots, M$. In each period $t = 1, \dots, T$, each consumer i has a shopping occasion, that is, visits the store, with probability ρ_i , which is exogenous, consumer-specific, and time-invariant.⁹ Utility is separable across categories, so a consumer’s choices in different categories are independent conditional on their preferences.¹⁰ We describe demand in a single category, omitting the category index. Consumers choose among J products available, indexed by $j = 1, \dots, J$, or the outside option (purchasing nothing), denoted by $j = 0$.

We assume that consumers have heterogeneous category tastes and price sensitivities. The indirect utility of consumer i buying product j in a given shopping occasion (suppressing the time index for notational simplicity) is

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

⁷We cannot model competition between supermarkets: our data cover a single supermarket, and no dataset records the purchases of all consumers in all competing supermarkets, as household scanner panels track only a sample of households.

⁸Our analysis assumes that the seller does not experiment with prices to accelerate learning, and that consumers do not withhold purchases to obtain lower prices. Section 5.6 studies some departures from these assumptions, where the seller or consumers try to influence learning.

⁹Store visit probabilities would respond to prices through the choice between competing supermarkets, a margin our single-supermarket data preclude.

¹⁰This abstracts from cross-category complementarities, such as between chips and soda (Ershov et al., 2025). Types may still be correlated across categories. The model does not restrict this correlation, and the estimation recovers the distribution of types category by category.

where δ_j is the mean utility of product j , which we interpret as its quality, capturing everything about the product that shifts all consumers' utility equally, β_i is consumer i 's taste for the product category, and $\alpha_i < 0$ is the price sensitivity. Prices may differ across consumers: under uniform pricing, $p_{ij} = p_j$ for all i ; under personalized pricing, they depend on the consumer's purchase history.

A consumer's type is defined as $\Theta_i = (\alpha_i, \beta_i, \rho_i)$, comprising the price sensitivity, category taste, and shopping occasion probability. The distribution of types in the population is characterized by the cumulative distribution function $F_\Theta(\Theta_i)$. We denote by $\theta_i = (\alpha_i, \beta_i)$ the preference parameters governing product choice. Types are category-specific, except for the shopping occasion probability ρ_i , which is common to all categories. We impose no restriction on how types correlate across categories.

Market shares. We make the standard assumption that the ϵ_{ij} are independent and identically distributed across individuals, products, and shopping occasions, and follow an extreme value Type I distribution. The utility of the outside option is normalized to $u_{i0} = \epsilon_{i0}$. Under these assumptions, the probability that a consumer with type Θ_i chooses product j , conditional on a shopping occasion, is

$$\mathbb{P}_j(\theta_i) = \frac{\exp(\delta_j + \beta_i + \alpha_i p_{ij})}{1 + \sum_{j'=1}^J \exp(\delta_{j'} + \beta_i + \alpha_i p_{ij'})}.$$

The joint probability that a consumer with type Θ_i has a shopping occasion and chooses product j is $s_j(\Theta_i) = \mathbb{P}_j(\theta_i) \cdot \rho_i$. Finally, the overall market share of product j is

$$s_j = \int \mathbb{P}_j(\theta_i) \cdot \rho_i dF_\Theta(\Theta_i). \quad (1)$$

Consumer surplus. The consumer surplus is the expected maximum utility, summed over consumers,

$$CS = -M \cdot \int \frac{\rho_i}{\alpha_i} \cdot \ln \left(1 + \sum_{j=1}^J \exp(\delta_j + \beta_i + \alpha_i p_{ij}) \right) dF_\Theta(\Theta_i), \quad (2)$$

where the surplus of a consumer without a shopping occasion is normalized to zero.

2.2 Supply and uniform pricing

Our model assumes that the supermarket sets retail prices and that manufacturers keep wholesale prices fixed regardless of the supermarket's pricing strategy.¹¹ The supermarket is therefore a multi-product monopolist facing exogenous marginal costs. Uniform pricing is the strategy the supermarket practices in the data and the baseline against which we measure the gains from personalization.

Expected profits under uniform pricing are

$$\pi^U = M \cdot \sum_{j=1}^J (p_j - c_j) s_j,$$

where c_j is the marginal cost of product j . The seller sets prices such that the following first-order conditions are satisfied,

$$\mathbf{s} + \mathbf{\Omega}(\mathbf{p} - \mathbf{c}) = 0, \quad (3)$$

where $\mathbf{s}$, $\mathbf{p}$, and $\mathbf{c}$ are the vectors of market shares (equation 1), prices, and marginal costs. $\mathbf{\Omega}$ is the matrix of market share derivatives, with element $\Omega_{jj'} = \partial s_{j'} / \partial p_j$:

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

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

2.3 Personalized pricing

Seller's learning from purchase histories. The seller observes each consumer's shopping occasions and, when they occur, the product chosen. The sequence of these choices over the τ periods of the observation phase is the consumer's purchase history, $\mathbf{h}_{\ell} = \{(Y_{\ell 01}, \dots, Y_{\ell J1}), \dots, (Y_{\ell 0\tau}, \dots, Y_{\ell J\tau})\}$, where $Y_{\ell jt}$ equals 1 if product j is purchased in period t and 0 otherwise.¹² We index histories by $\ell = 1, \dots, L$. The number of consumers with

¹¹Berto Villas-Boas (2007) and Bonnet and Dubois (2010) estimate richer models of vertical contracting between manufacturers and retailers.

¹²If a shopping occasion occurs in period t , exactly one option is chosen: $\sum_{j=0}^J Y_{\ell jt} = 1$. If no shopping occasion occurs, $\sum_{j=0}^J Y_{\ell jt} = 0$.

history $\mathbf{h}_\ell$ is $M_{\mathbf{h}_\ell}$, so that $M = \sum_{\ell=1}^L M_{\mathbf{h}_\ell}$.¹³

We assume that the seller updates beliefs about each consumer's type by applying Bayes's rule, combining the prior population distribution with the observed purchase history.¹⁴ Bayesian updating is the natural and optimal inference rule here: in statistical decision theory, no procedure can dominate it, and any rule that ignores prior information can always be matched or improved upon by a Bayesian one (Wald, 1950 and Berger, 1985). This optimality requires that the seller knows the population distribution of types, and that preferences do not change over time. The seller also learns category by category, from the purchase history in that category alone.¹⁵ Following Bayes's rule, we can write

$$f_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell) = \frac{f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i) \cdot f_{\Theta}(\Theta_i)}{f_{\mathbf{h}_\ell}(\mathbf{h}_\ell)}. \quad (4)$$

This expression has a natural interpretation: the posterior probability of type Θ_i exceeds its prior probability when the realized history is more likely conditional on type Θ_i than on average.

We now decompose this expression. Because demand shocks are independent across shopping occasions, the probability of observing history $\mathbf{h}_\ell$ conditional on Θ_i is the product of the period-by-period probabilities:

$$f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i) = \prod_{t=1}^{\tau} \left[\prod_{j=0}^J \mathbb{P}_{jt}(\theta_i)^{Y_{\ell jt}} \right] \rho_i^{\sum_{j=0}^J Y_{\ell jt}} (1 - \rho_i)^{1 - \sum_{j=0}^J Y_{\ell jt}}, \quad (5)$$

where the time subscript on $\mathbb{P}_{jt}$ allows prices to vary over the observation phase. The unconditional probability of observing purchase history $\mathbf{h}_\ell$ is

$$f_{\mathbf{h}_\ell}(\mathbf{h}_\ell) = \int f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i) dF_{\Theta}(\Theta_i). \quad (6)$$

Equation (5) distinguishes choosing the outside option from not having a shopping occasion. In the first case, the seller observes a choice at known prices and updates its beliefs about the

¹³The number of possible histories grows exponentially in τ . The posterior, however, depends on the history only through a low-dimensional sufficient statistic, whose distribution we compute by recursion, as explained in Appendix C.1.

¹⁴The seller in our model learns about consumers within a known demand model, which differs from the pricing algorithms of Calvano et al. (2020) and Asker et al. (2024), which learn demand itself by trial and error, varying prices to discover the returns to each.

¹⁵We relax two of these assumptions at the end of Section 5.4. First, the seller learns across categories and pools the histories of two pairs of frequently co-purchased categories. Second, the seller does not know the population distribution of types and updates from a misspecified prior.

consumer's preference parameters directly. In the second case, it observes no choice, but the absence of a visit is informative about ρ_i , and through the correlation between ρ_i and θ_i in the population, about the preference parameters as well.

Optimal prices. After the observation phase, the seller price discriminates: consumers with purchase history $\mathbf{h}_\ell$ face the price vector $\mathbf{p}(\mathbf{h}_\ell)$. Expected profits are

$$\pi^H = \sum_{\ell=1}^L M_{\mathbf{h}_\ell} \sum_{j=1}^J (p_j(\mathbf{h}_\ell) - c_j) s_j(\mathbf{h}_\ell),$$

where $s_j(\mathbf{h}_\ell) = \int \mathbb{P}_j(\theta_i) \rho_i dF_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell)$ is the expected market share of product j among consumers with purchase history $\mathbf{h}_\ell$. For each purchase history $\mathbf{h}_\ell$, the optimal personalized prices satisfy the system of J first-order conditions:

$$\mathbf{s}(\mathbf{h}_\ell) + \mathbf{\Omega}(\mathbf{h}_\ell) (\mathbf{p}(\mathbf{h}_\ell) - \mathbf{c}) = 0. \quad (7)$$

The vector $\mathbf{s}(\mathbf{h}_\ell)$ collects the history-specific market shares, and $\mathbf{\Omega}(\mathbf{h}_\ell)$ is the matrix of market share derivatives with the integrals taken against the posterior distribution $F_{\Theta|\mathbf{h}_\ell}$. The seller solves this system independently for each of the L purchase histories.

2.4 Perfect price discrimination

As a benchmark, we take the extreme case where the seller perfectly knows the type of each consumer and conditions its price on it. This case bounds what learning from purchase histories can achieve. Expected profits are

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

Since ρ_i multiplies every term of a consumer's expected profit, it drops out of the first-order conditions, and the optimal prices depend on the preference parameters θ_i only:

$$\mathbb{P}(\theta_i) + \mathbf{\Omega}(\theta_i)(\mathbf{p}(\theta_i) - \mathbf{c}) = 0, \quad (8)$$

where $\mathbb{P}(\theta_i)$ is the vector of choice probabilities and $\mathbf{\Omega}(\theta_i)$ is the matrix of their price derivatives, with element $\Omega_{jj'}(\theta_i) = \partial \mathbb{P}_{j'}(\theta_i) / \partial p_j$.

3 Data

3.1 Raw data

We estimate the preference parameters and marginal costs using proprietary transaction data from DecaData.¹⁶ The full database contains daily point-of-sale transactions from 853 grocery stores across 13 U.S. states over several years. Our analysis uses a single store, located near the median in terms of revenues, and calendar year 2018. The data record individual transactions: each observation represents the purchase of one product, identified by its universal product code (UPC), within a customer’s checkout basket, together with the price paid, the quantity purchased, and the product’s manufacturer and classification into categories. The data also include each store’s daily shipment records: for every delivery, we observe the date, the product UPC, the quantity delivered, and the total invoice cost. We use the records of 2017 and 2018, deriving a unit wholesale price by matching each sale to the most recent prior delivery of that product. We use these wholesale prices twice: to construct the margins targeted in the estimation, and to validate the marginal costs recovered from the model.

Crucially for our analysis, the data include consumer identifiers for customers enrolled in the stores’ loyalty programs, allowing us to track individual purchase histories over time.¹⁷ This is a key advantage over store-level scanner data, which record only aggregate sales, and over household panels, which track only a small sample of any store’s customers. The consumer identifiers link purchases along two dimensions, and we use both. Over time, repeated purchases identify the distribution of latent consumer types in each category. Across categories, joint purchases reveal how each consumer combines category types, which determines how personalized prices aggregate over grocery baskets. The drawback of the data is that they include neither consumer demographics, nor product characteristics, nor store characteristics beyond each store’s state. The absence of demographics matters little for our purpose: in the two studies closest to ours, purchase histories raise the profits from personalized pricing far more than demographics do (Rossi et al., 1996, Smith et al., 2023).

¹⁶Source: <https://decadata.io/data>.

¹⁷Consumer identifiers are store-specific. We cannot track purchases across different stores.

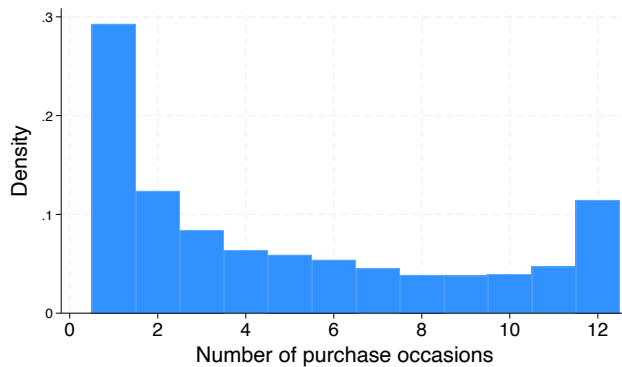
3.2 Sample construction

We restrict attention to a single store since analyzing several would replicate the same learning and pricing exercise without additional insight. To avoid the idiosyncrasies of very large or very small stores, we select a store near the median of the revenue distribution; it generated \$13.1 million in sales in 2018, placing it in the 52nd percentile of the 853 stores.

We limit our analysis to the 20,748 consumers with loyalty cards who made at least one purchase at this store in 2018. These consumers spend on average \$42 per month across all categories (\$102 in months with at least one store visit), for a total of \$10.4 million in annual sales. These purchases represent 79.6% of the store’s total revenue and 84.9% of all transactions, suggesting that our sample captures the bulk of the store’s regular customer base. We focus on 24 of the largest product categories spanning six departments, representing approximately 20% of the supermarket’s total sales. Additional details on the construction of the sample are relegated to Appendix [A.1](#).

We aggregate purchases to the consumer-month level. A consumer with at least one purchase in a month, in any product category (including the excluded ones), has a shopping occasion in that month. Otherwise, the consumer has no shopping occasion in that month. [Figure 1](#) presents the histogram of the number of shopping occasions over the twelve months. The distribution is U-shaped: both low and high numbers of shopping occasions are common, suggesting important heterogeneity in shopping occasion probabilities. By construction, consumers must visit the store at least once to be included in the sample, so the distribution

Figure 1: Distribution of shopping occasions over 12 months



Notes: Number of months with at least one purchase at the store in 2018, over the 20,748 sampled consumers.

is truncated at one.

3.3 Evidence of uniform pricing

We use the transaction-level data to ask whether the store already charges different prices to different consumers. Consumers who buy the same UPC on the same day pay the same price in 92.6% of UPC-days. We find that cardholders pay less than shoppers without a loyalty card, and unit prices decrease with the quantity purchased. Once we restrict the sample to the single-unit purchases of cardholders, 96.3% of UPC-days have a single price.

In this restricted sample, we regress the price paid on the consumer’s shopping frequency and, separately, on the consumer’s annual expenditure at the store, with UPC-day fixed effects. Prices would vary with these variables if the store conditioned prices on purchase histories. Both effects are tiny, less than half a cent for an additional month of shopping or for a doubling of annual expenditure. Appendix A.2 reports these results in detail. The store offers quantity discounts and discounts for cardholders, but it does not condition prices on purchase histories. The personalized pricing we study is therefore a departure from the store’s current pricing rather than a refinement. The evidence also supports the assumption of uniform pricing under which we recover marginal costs.

4 Estimation, Identification, and Estimation Results

4.1 Estimation framework

We seek to identify the joint distribution of consumer types F_Θ and the mean utilities δ_{jt} , which vary by period in the estimation. A period is a month, so $T = 12$. We collect the period- t mean utilities in the vector $\boldsymbol{\delta}_t = (\delta_{1t}, \dots, \delta_{Jt})$. The utility of consumer i buying product j in period t is $u_{ijt} = \delta_{jt} + \beta_i + \alpha_i p_{jt} + \epsilon_{ijt}$. Following Fox et al. (2011), we approximate the distribution of consumer types by a discrete distribution with D fixed points of support.¹⁸ Therefore, identifying F_Θ amounts to estimating a probability mass function $\{\phi_d\}_{d=1,\dots,D}$ and the points of support $\{\Theta_d\}_{d=1,\dots,D}$. The discrete approximation is appealing in our context for two reasons. First, it allows for rich and flexible consumer heterogeneity that produces arbitrary correlations among α , β , and ρ . Second, it keeps the seller’s Bayesian updating

¹⁸The discrete types are support points of this approximation, not a claim that the population contains finitely many types.

tractable: with D types, the posterior distribution after any purchase history is a vector of D probabilities.

We consider the likelihood of observing the purchase histories in the data, treating each consumer's entire history as a single observation. We observe a sample of M individuals and their respective purchase histories $\mathbf{h}_i = \{Y_{ijt}\}_{j=0,\dots,J;t=1,\dots,T}$. Consumer i 's history realizes one of the patterns $\mathbf{h}_\ell$ of Section 2. The contribution of consumer i to the likelihood is the unconditional probability of observing history $\mathbf{h}_i$,

$$\Pr(\mathbf{h}_i) = \sum_{d=1}^D \phi_d \prod_{t=1}^T \left[\prod_{j=0}^J \mathbb{P}_{jt}(\theta_d, \boldsymbol{\delta}_t)^{Y_{ijt}} \right] \rho_d^{\sum_{j=0}^J Y_{ijt}} (1 - \rho_d)^{1 - \sum_{j=0}^J Y_{ijt}}, \quad (9)$$

as per equation (6), evaluated at the discrete type distribution. The choice probabilities

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

give the probability that a type- d consumer chooses product j conditional on having a shopping occasion in period t .

To reduce the dimensionality of the estimation problem, we include a set of constraints based on the market share equation (1) to identify $\boldsymbol{\delta}_t$: for every product j and period t , we impose that

$$\sum_{d=1}^D \phi_d \mathbb{P}_{jt}(\theta_d, \boldsymbol{\delta}_t) \rho_d = \mathfrak{s}_{jt}, \quad (10)$$

where $\mathfrak{s}_{jt}$ is the observed aggregate market share. Given the type distribution, this system defines a unique vector of mean utilities $\boldsymbol{\delta}_t$ in each period, as shown by [Berry \(1994\)](#).

4.2 Estimation procedure

Our estimation method combines elements from the nonparametric estimator described in [Fox et al. \(2011\)](#) and [Fox et al. \(2016\)](#) with the MPEC formulation of the BLP model (see [Dubé et al., 2012](#) and [Grieco et al., 2025](#)), which relies on a constrained optimization to recover the $\{\boldsymbol{\delta}_t\}_{t=1,\dots,T}$. The estimation proceeds in four stages. First, we estimate the marginal distribution of the shopping occasion probability ρ , which the subsequent stages take as given. Second, we estimate the supports of the preference parameters θ in each product category, fixing ρ at its estimated mean to reduce dimensionality. The likelihood is invariant to a common shift of the price coefficients or the category tastes, compensated through the mean

utilities. This stage thus identifies the widths of the supports but not their locations, which are fixed at arbitrary anchors. Third, we construct a discrete grid spanning the estimated supports and estimate the type probabilities $\{\phi_d\}_{d=1,\dots,D}$ and the mean utilities $\{\delta_t\}_{t=1,\dots,T}$ by constrained maximum likelihood. Fourth, we set the location of the price coefficient support using the supply side of the model. Marginal costs recovered by inverting the seller’s first-order conditions for uniform pricing, equation (3), as $\mathbf{c} = \mathbf{p} + \mathbf{\Omega}^{-1}\mathbf{s}$, must reproduce the observed wholesale margins on average. The location of the category taste support is a free normalization, which we resolve by normalizing the category taste to have mean zero.

First stage: Estimation of the shopping occasion distribution. We approximate the heterogeneity in shopping occasions with a finite mixture: the population consists of U latent classes with shopping occasion probabilities $\rho_1, \dots, \rho_U$. A consumer in class u has a shopping occasion each month with probability ρ_u , so the number of shopping occasions n_i observed over the T months follows a Binomial distribution:

$$n_i \sim \text{Binomial}(T, \rho_u).$$

We estimate the shopping occasion probabilities $\{\rho_u\}_{u=1,\dots,U}$ and the class probabilities $\{\psi_u\}_{u=1,\dots,U}$, where $\psi_u = \Pr(\rho = \rho_u)$, by fitting the finite mixture to the observed distribution of the yearly occasion counts n_i across consumers. The sample contains only cardholders with at least one shopping occasion during the observation window, so we maximize the likelihood of the observed counts conditional on this event: the mixture probability of each positive count is divided by the probability of at least one occasion, $1 - \sum_{u=1}^U \psi_u (1 - \rho_u)^T$. This zero-truncated likelihood recovers the population class probabilities, including the share of consumers who never visit the store during the year, from the positive counts alone. In the estimation, ρ_u is parameterized with a logit link function to ensure it lies in $(0, 1)$.

Second stage: Estimation of preference parameter supports. We follow the location and scale model of Fox et al. (2011) to estimate the supports of the preference parameters: α ranges over $(a_1, a_1 + a_2)$ and β over $(b_1, b_1 + b_2)$. In this stage, α and β follow independent uniform distributions over their supports. To further simplify the problem, we fix ρ for all consumers at $\bar{\rho}$, the average shopping occasion frequency observed in the sample.¹⁹ This

¹⁹Fixing ρ at $\bar{\rho}$ is equivalent to assuming that (α_i, β_i) is independent of ρ_i in this stage. The shopping occasion terms then factor out of the likelihood as a constant, and ρ enters the market share constraint only through its mean $\bar{\rho}$.

assumption serves only to estimate the support, and the third stage restores the integration over the marginal distribution of ρ and allows for unrestricted correlations between α , β , and ρ .

The choice probabilities depend on the parameters only through the sum $\delta_{jt} + \beta_i + \alpha_i p_{jt}$, so shifting every α_i by a constant κ while adjusting the mean utilities to $\delta_{jt} - \kappa p_{jt}$, or shifting every β_i by κ while adjusting them to $\delta_{jt} - \kappa$, leaves all choice probabilities unchanged. The locations of the supports are therefore not identified. To circumvent this issue at this stage of the estimation routine, we anchor the supports provisionally by fixing their lower bounds, $a_1 = 0$ and $b_1 = 0$, and focus on estimating the width of the supports, a_2 and b_2 . The location of the price coefficient support is set in the fourth stage using the supply side of the model, and the category taste is renormalized to mean zero once the type distribution is estimated.

We solve the following estimation problem,

$$\begin{aligned} \max_{a_2, b_2, \{\delta_t\}_{t=1, \dots, T}} \quad & \sum_{i=1}^M \ln \left(\frac{1}{R} \sum_{r=1}^R \prod_{t=1}^T \prod_{j=0}^J \mathbb{P}_{jt}(\theta_r, \delta_t)^{Y_{ijt}} \right), \\ \text{subject to} \quad & \bar{\rho} \cdot \frac{1}{R} \sum_{r=1}^R \mathbb{P}_{jt}(\theta_r, \delta_t) = \mathfrak{z}_{jt}, \quad \forall j, t, \\ & \alpha_r = a_2 \eta_{r1}, \quad \beta_r = b_2 \eta_{r2}, \quad r = 1, \dots, R, \end{aligned} \tag{11}$$

where η_{r1} and η_{r2} are Halton draws used to approximate the integral. We use $R = 50$ draws. The widths are estimated from several starting values, as the objective admits local optima.

Third stage: Estimation of the distribution of preference parameters. We construct a grid of consumer types by combining the U values of ρ with 40 pairs of (α, β) drawn from their estimated supports using Halton sequences, yielding an initial grid of 40 U types. The grid points $(\alpha_d, \beta_d, \rho_d)$ are held fixed during the estimation. We estimate the type probabilities and the mean utilities separately for each category by maximum likelihood, accounting for the selection of observed histories: a consumer without any shopping occasion over the T months cannot appear in the sample, so the likelihood of an observed history is conditional on

having at least one occasion. With contributions $\Pr(\mathbf{h}_i)$ given in equation (9), we solve

$$\begin{aligned}
& \max_{\{\phi_d\}_{d=1,\dots,D}, \{\delta_t\}_{t=1,\dots,T}} \sum_{i=1}^M \ln \left[\frac{\Pr(\mathbf{h}_i)}{1 - \sum_{d=1}^D \phi_d (1 - \rho_d)^T} \right] \\
& \text{subject to} \quad \sum_{d=1}^D \phi_d \mathbb{P}_{jt}(\theta_d, \delta_t) \rho_d = \varsigma_{jt}, \quad \forall j, t, \\
& \sum_{d=1}^D \phi_d = 1, \quad \phi_d \geq 0, \\
& \sum_{d=1}^D \phi_d \mathbb{1}(\rho_d = \rho_u) = \psi_u, \quad u = 1, \dots, U.
\end{aligned} \tag{12}$$

The last set of constraints ensures that the marginal distribution of ρ matches its population distribution from the first stage, $\{\psi_u\}_{u=1,\dots,U}$. The denominator of the objective is the probability of at least one shopping occasion, which makes the sampling process explicit, since only consumers with at least one occasion can appear in the data. Given the last set of constraints, this denominator equals the constant $1 - \sum_{u=1}^U \psi_u (1 - \rho_u)^T$, so the conditioning shifts the conditional log-likelihood by a constant.²⁰ Consistent with the population interpretation of $\{\phi_d\}_{d=1,\dots,D}$, the market size used to compute the observed shares ς_{jt} is the population of cardholders, obtained by dividing the number of sampled consumers by the first-stage probability of at least one occasion. Consumers who never visit the store are part of the market even though they do not appear in our estimation sample.

The estimated distribution typically spreads small amounts of mass over many neighboring grid points across which the likelihood is nearly flat. We therefore consolidate the types that the likelihood does not distinguish from each other, which lowers the dimension of the posterior distributions computed in the counterfactuals.²¹

Fourth stage: The location of the price coefficient support. It remains to estimate the location of the price coefficient's support, which after the second stage normalization

²⁰The selection is fully controlled by the first-stage estimate of the distribution of ρ and the estimated type probabilities $\{\phi_d\}_{d=1,\dots,D}$ coincide with those obtained from maximizing the unconditional likelihood.

²¹The taste pairs (α, β) form a common grid shared by the shopping occasion frequency classes. We iteratively replace a pair of taste points by a single point located at their probability weighted centroid, which inherits their probability mass within each frequency class. At each step we merge the pair with the smallest value of Ward's criterion, $\frac{\phi_d \phi_{d'}}{\phi_d + \phi_{d'}} x_{dd'}^2$ with $x_{dd'}^2 = \left(\frac{\alpha_d - \alpha_{d'}}{a_2} \right)^2 + \left(\frac{\beta_d - \beta_{d'}}{b_2} \right)^2$, and recompute the exact log-likelihood of the consolidated representation. We stop when the next merger would reduce the total log-likelihood by more than one log point relative to the full estimate.

coincides with the lower bound a_1 . Demand at the observed prices is invariant to the location parameter, but its derivatives with respect to price are not: more negative price coefficients imply more elastic demand and smaller equilibrium margins. Inverting equation (3) therefore yields marginal costs that depend on the location parameter a_1 ,

$$\mathbf{c}_t(a_1) = \mathbf{p}_t + \mathbf{\Omega}_t(a_1)^{-1} \mathbf{s}_t, \quad (13)$$

where $\mathbf{\Omega}_t(a_1)$ is the matrix of market share derivatives with respect to price evaluated at the price coefficients $\alpha_d + a_1$ and the correspondingly adjusted mean utilities $\delta_{jt} - a_1 p_{jt}$. We use the observed wholesale margins m_{jt} , defined as the difference between the retail price and the wholesale price in the invoice data, to identify the location of the price coefficient a_1 . The underlying assumption is that the marginal cost of selling a product is its wholesale price and the supermarket's retail costs, which include handling and stocking, are negligible.

We estimate a_1 so that implied and observed margins match on average, for all products except the composite good, weighting by sales:

$$\sum_t \sum_{j \in \mathcal{J}} \omega_{jt} [p_{jt} - c_{jt}(a_1)] = \sum_t \sum_{j \in \mathcal{J}} \omega_{jt} m_{jt}, \quad \text{with } \omega_{jt} = \frac{\delta_{jt}}{\sum_{t'} \sum_{j' \in \mathcal{J}} \delta_{j't'}}. \quad (14)$$

The set $\mathcal{J}$ collects the product and period cells of all manufacturers with an observed wholesale margin, excluding the composite good. The composite good aggregates many small brands, so its measured margin is an average over brands with separate prices and invoice costs, which we consider less reliable than a single brand's margin. Its marginal costs are still recovered from equation (13). This is one equation with one unknown, so we solve this nonlinear equation directly. The final estimates are the shifted type grid and the correspondingly adjusted $\{\delta_t\}_{t=1, \dots, T}$; the category taste keeps its mean zero normalization.

This stage also backs out the marginal cost of every product in every period, evaluated at the estimated location of α . We emphasize that only the average wholesale margin is matched by construction, and the product-level recovered costs are not fitted to the observed margins.

4.3 Identification

Our identification strategy builds on the nonparametric identification results in [Fox et al. \(2016\)](#) for finite mixture models of multinomial choice, adapted to our setting in two respects:

consumers are observed repeatedly across shopping occasions, and market share constraints identify the mean utilities.

Consumers with different types Θ_d generate systematically different purchase histories, and this variation across consumers identifies the type probabilities $\{\phi_d\}_{d=1,\dots,D}$, which, given the fixed support points, determine the distribution of types. Each dimension of the consumer type $(\alpha_d, \beta_d, \rho_d)$ is identified by a distinct feature of the purchase histories. Because prices are uniform, all consumers face the same sequence of prices, so differences in purchase behavior across consumers reflect differences in preferences rather than in the choice environment. The dispersion in price sensitivity is identified by how differently consumers respond to the same price changes: more price-sensitive consumers concentrate their purchases in low-price periods and low-price products. The category taste β_d is common to all products, so it cancels from comparisons among products and matters only for the choice between purchasing in the category and the outside option. Its heterogeneity is identified by the variation across consumers in how often they purchase in the category rather than choosing the outside option, conditional on prices. The parameter ρ_d is separately identified from the preference parameters by the observed frequency of shopping occasions, regardless of what is purchased. The purchase histories identify the dispersions of α and β but not their locations. As shown in Section 4.2, a common shift of either coefficient is absorbed by the mean utilities, leaving choice probabilities unchanged.

The mean utilities $\{\delta_t\}_{t=1,\dots,T}$ are identified by the market share constraints in equation (10), which, given the type distribution, uniquely pin down the mean utilities that rationalize the observed market shares. Their location inherits the location normalizations of the preference coefficients and is resolved in the fourth stage.

Finally, the location of the price coefficients' support and the marginal costs are identified jointly by the supply side.²² We assume that the observed prices are the optimal uniform prices of a monopolistic, multi-product seller taking wholesale prices as given, so that the first-order conditions in equation (3) can be inverted for marginal costs, provided the matrix of market share derivatives Ω is invertible. The estimated marginal costs depend on the location

²²In our context, the choice data alone cannot identify the location of the price coefficients' support. Traditionally, one would define $\delta_{jt} = \bar{\alpha}p_{jt} + \xi_{jt}$, where ξ_{jt} is an unobserved residual, and use an instrumental variable regression to recover the mean price sensitivity $\bar{\alpha}$. Since we do not have access to product characteristics, we could not construct the usual BLP instruments or differentiation instruments (see [Berry et al., 1995](#) and [Gandhi and Houde, 2026](#)).

of the price coefficients through the price sensitivity of demand: more price-sensitive demand implies smaller equilibrium margins. The location of the distribution of price sensitivities is identified as the value at which the recovered margins reproduce the observed margins on average, according to equation (14).

This identification strategy also clarifies why the endogeneity of prices, the central concern of demand estimation with market-level data, has limited scope in our setting. The period-specific mean utilities absorb all aggregate demand variation, so any correlation between prices and market-level demand shocks is captured by δ_{jt} . Therefore, since prices are uniform across consumers, the cross-consumer variation that identifies the heterogeneity parameters cannot be driven by prices responding to individual preferences. The level of the price sensitivity is the one object that price endogeneity would ordinarily bias, and precisely the one we do not estimate from choice data. It is identified from observed margins under the conduct assumption, so no instrument for prices is required.

4.4 Estimation results

4.4.1 Shopping occasion distribution

We use $U = 3$ latent classes, which parsimoniously capture the three salient shopping frequencies in the data. The estimates, reported in Table B.1, reveal that the largest group (58.7%) consists of occasional shoppers with a 9.4% shopping probability in any given month. Regular shoppers comprise 24.7% of consumers and shop with 50.3% probability, while frequent shoppers (16.7%) shop with 94.2% probability per month. The estimates imply that 18.0% of cardholders make no visits during the twelve-month window.

4.4.2 Preference parameters

Table 1 summarizes the patterns of preference heterogeneity. For each category, we report the mean and standard deviation of price sensitivity, α_d , and the standard deviation of category taste, β_d (normalized to a zero mean). The final column quantifies product differentiation in dollars: the standard deviation of month-averaged mean utilities divided by the absolute mean price sensitivity, capturing the spread of price-equivalent qualities.

The results reveal substantial heterogeneity both within and across categories. Within categories, the standard deviation of the price coefficient ranges from 9% (cereal) to 38% (wine) of its mean, with the two alcohol categories displaying the highest dispersion. The

standard deviation of category tastes ranges from 0.76 (crackers) to 2.2 (bleach). The shopping occasion probability is correlated with preferences. Frequent shoppers have stronger category tastes in 22 of the 24 categories and are less price sensitive in 18. [Table B.2](#) reports the estimated supports and the correlations between α , β and ρ for every category.

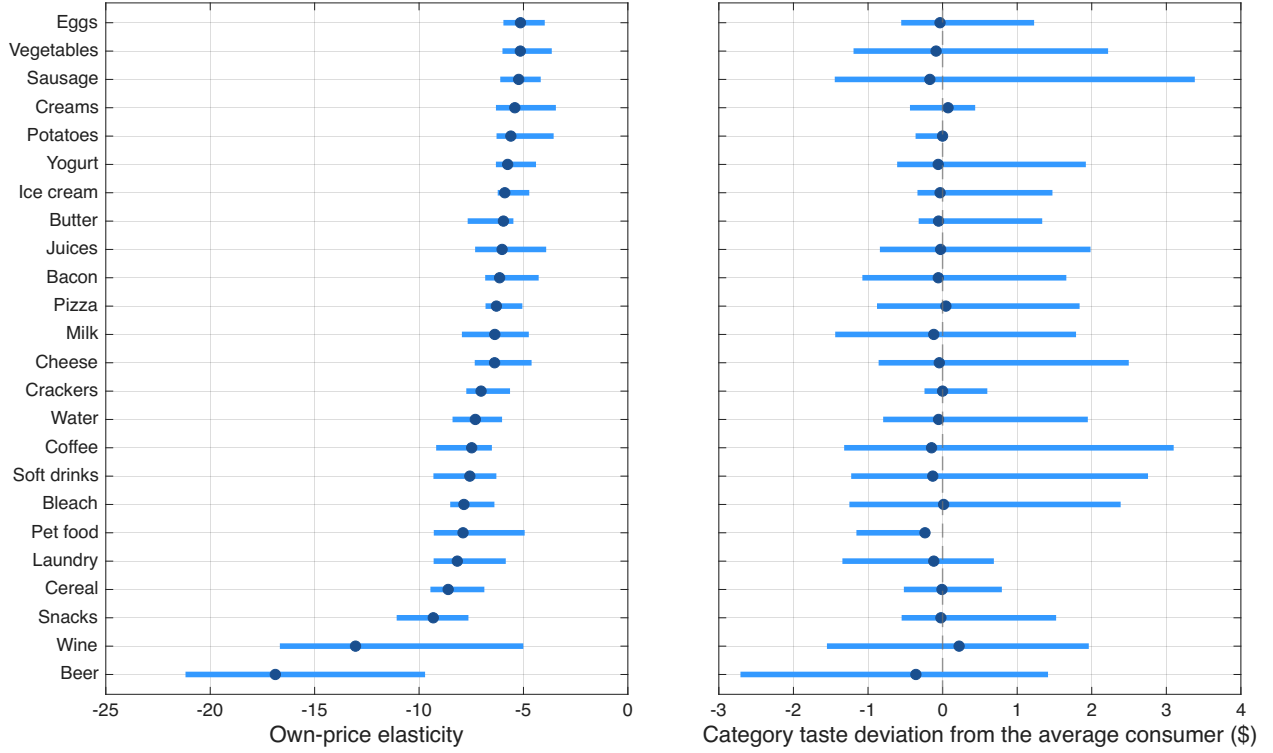
[Figure 2](#) displays, for each category, the means and the 10th to 90th percentile ranges of the own price elasticities and of the category tastes, converted to dollars, across consumer types. Mean own price elasticities range from -5.1 (fresh eggs) to -16.9 (beer). The two alcohol categories are far more elastic than the others, which lie between -5.1 and -9.3 (snacks). These magnitudes are in line with [Atalay et al. \(2026\)](#) for bottled water and carbonated soft drinks. They report average elasticities of -6.8 and -9.6 at the start of their sample, bracketing our estimates of -7.3 and -7.6. Our demand estimates are however more elastic than those of the other at-scale studies of grocery demand. [Döpper et al. \(2025\)](#) report a mean across category-years of the median own price elasticity of -2.6, while [Bhattacharya et al. \(2026\)](#) estimate a median across product markets of -2.9. One possible explanation is the market definition: both studies estimate demand for brands in regional markets that pool many retailers, whereas we estimate demand at a single store, whose consumers can also substitute toward other retailers. [Figure B.1](#) and [Figure B.2](#) show the distributions of own price elasticities and category tastes (in dollars) across consumer types in each category.

Table 1: Summary of preference parameters across categories

Category	α_d		β_d	Quality
	Mean	Std. dev.	Std. dev.	disp. (\$)
Alcohol				
Beer	-0.998 (0.073)	0.295 (0.037)	1.504 (0.441)	2.121
Wine	-0.750 (0.105)	0.287 (0.042)	0.929 (0.344)	1.184
Dairy				
Butter/margarine	-1.394 (0.051)	0.229 (0.024)	1.080 (0.108)	2.640
Cheese	-0.850 (0.012)	0.107 (0.008)	1.063 (0.094)	1.543
Creams	-1.073 (0.098)	0.244 (0.032)	1.292 (0.253)	1.584
Fresh eggs	-1.511 (0.027)	0.211 (0.018)	1.038 (0.113)	0.074
Refr. juices	-1.034 (0.020)	0.231 (0.009)	1.334 (0.070)	1.516
Milk	-1.184 (0.051)	0.203 (0.023)	1.369 (0.138)	1.436
Yogurt	-1.097 (0.058)	0.183 (0.028)	1.073 (0.115)	0.778
Frozen food				
Ice cream	-0.888 (0.018)	0.110 (0.013)	0.890 (0.047)	1.401
Fz. pizza	-0.775 (0.117)	0.096 (0.054)	1.252 (0.237)	1.959
Fz. potatoes	-1.222 (0.028)	0.217 (0.013)	0.826 (0.055)	0.672
Fz. vegetables	-0.866 (0.021)	0.169 (0.014)	0.870 (0.079)	0.627
Grocery				
Cereal	-1.205 (0.016)	0.110 (0.011)	0.945 (0.075)	2.336
Coffee	-0.696 (0.037)	0.112 (0.018)	1.328 (0.178)	15.445
Crackers	-1.405 (0.016)	0.173 (0.008)	0.755 (0.059)	1.572
Snacks	-1.239 (0.039)	0.134 (0.008)	0.906 (0.071)	2.431
Soft drinks	-0.675 (0.018)	0.104 (0.008)	1.452 (0.076)	3.909
Water	-0.850 (0.024)	0.109 (0.015)	1.249 (0.121)	7.312
Packaged meat				
Bacon	-0.616 (0.019)	0.097 (0.012)	0.778 (0.079)	1.612
Dinner sausage	-0.564 (0.016)	0.106 (0.010)	1.202 (0.066)	4.584
Taxable grocery				
Bleach	-1.571 (0.057)	0.157 (0.022)	2.196 (0.417)	5.454
Laundry detergent	-0.915 (0.032)	0.154 (0.016)	1.171 (0.111)	4.471
Pet food	-0.608 (0.093)	0.163 (0.031)	1.206 (0.356)	2.456

Notes: Statistics are weighted by the estimated type probabilities ϕ_d . The category taste β_d is normalized to have mean zero in each category. Quality dispersion is the standard deviation across products of $\bar{\delta}_j/|\bar{\alpha}|$, where $\bar{\delta}_j$ is the month average of the estimated mean utilities of product j and $\bar{\alpha}$ is the mean price coefficient. It measures the spread of price equivalent qualities in dollars. Bootstrap standard errors in parentheses, from 200 replications that resample the 20,748 cardholders with replacement and re-estimate the model. The standard errors use the replications that converge, between 135 and 162 in wine, frozen potatoes and frozen pizza and at least 192 in every other category.

Figure 2: Price elasticity and category taste deviation across categories



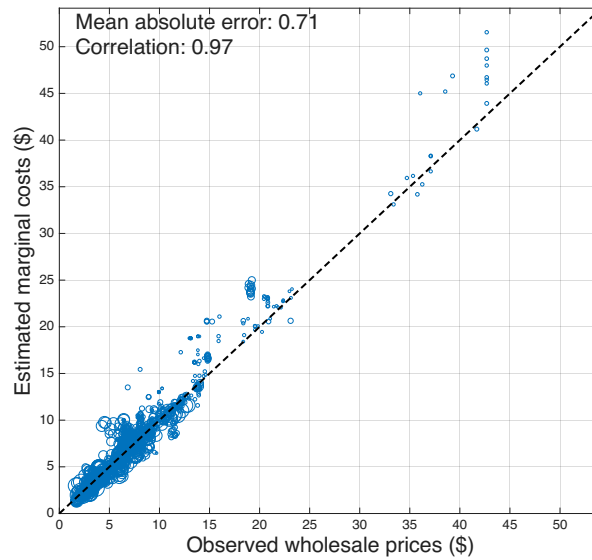
Notes: Each row summarizes the estimated own-price elasticities and category tastes. Bars span the 10th to 90th percentiles and dots mark the means, weighted by the estimated type probabilities ϕ_d . The left panel shows the type-level own-price elasticity, the sales-weighted average over products and months of $\alpha_d, p_{jt}(1 - \mathbb{P}_{jt}(\theta_d, \delta_t))$, evaluated at observed prices; the weights are common to all types so differences across types reflect preferences rather than purchase baskets. The right panel shows the type-level category taste in dollars, the deviation of β_d from the category mean divided by $|\alpha_d|$; the dashed vertical line at zero marks the mean category taste. Categories are sorted by mean own-price elasticity.

4.4.3 Marginal costs

We validate the model by comparing the estimated marginal costs with the wholesale prices observed in the invoice data. The comparison is informative even though the estimation matches wholesale margins on average. We fit a single sales-weighted average margin per category, so our product-level comparison is not a mechanical consequence of the estimation. The wholesale prices give us a check that studies recovering marginal costs from first-order conditions usually cannot perform.

Figure 3 presents a scatter plot of the estimated marginal costs and the observed wholesale prices, over all categories, with each product-month represented by a circle of size proportional to the product’s market share. The estimates lie close to the 45-degree line. Over the 1,320 product-months with an observed wholesale margin, the correlation between the estimated marginal costs and the observed wholesale prices is 0.974, and the share weighted mean absolute error is \$0.71. Excluding all composite products, whose margins aggregate over many small manufacturers, the correlation rises to 0.981 and the mean absolute error falls to \$0.67. Costs recovered from the supermarket’s pricing behavior thus align closely with wholesale prices observed in invoice data, consistent with the demand estimates and with our assumption that the supermarket sets optimal uniform prices.

Figure 3: Estimated marginal costs vs. observed wholesale prices



Notes: Each circle represents one product-month; its size is proportional to the product’s market share. Both axes are in dollars per standardized unit, and the dashed line is the 45 degree line.

5 Quantifying the Effects of Personalized Pricing

5.1 Setup

We now study how the supermarket learns about individual consumers from their purchase histories and uses that information to personalize prices. The supermarket observes each consumer’s shopping occasions and choices, knows the distribution of preferences in the population, prices, and product qualities, and conditions prices on the resulting purchase histories. To examine how the value of information evolves with data availability, we vary the length of the purchase histories used for personalization from one month to three years. We simulate purchase histories at each length from the estimated model and integrate over their probabilities of occurring. We implement personalized prices in the environment of one month of our data, so the personalization period has the product qualities and marginal costs of October 2018 at every history length. Computational details are provided in [Appendix C.1](#).

5.2 Information acquisition

We first characterize how the seller learns as it acquires longer purchase histories. For each preference parameter $\eta = \{\alpha, \beta, \rho\}$, we summarize learning by the proportional reduction in the variance of η from conditioning on a purchase history of length τ , averaged over all possible histories, that is,

$$\mathcal{I}_\eta(\tau) = 1 - \frac{\sum_{\ell=1}^L \sigma_{\eta|\mathbf{h}_\ell^\tau}^2 \cdot f_{\mathbf{h}_\ell}(\mathbf{h}_\ell^\tau)}{\sigma_\eta^2},$$

where $\sigma_\eta^2 = \sum_{d=1}^D (\eta_d - \mathbb{E}(\eta))^2 \phi_d$ is the variance of η in the population and $\sigma_{\eta|\mathbf{h}_\ell^\tau}^2 = \sum_{d=1}^D (\eta_d - \mathbb{E}(\eta | \mathbf{h}_\ell^\tau))^2 \phi_{d|\mathbf{h}_\ell^\tau}$ is the variance under the posterior induced by history $\mathbf{h}_\ell^\tau$. The index equals zero when the posterior variance of η equals its population variance after every history, so that histories are uninformative about η . It equals one when the posterior variance is zero after every history, so that the seller knows each consumer’s value of η . It increases with history length.²³ By the law of total variance, the population variance of η is the sum of the

²³Alternative learning indices based on the Kullback–Leibler divergence of the posterior from the prior, or on the reduction of its entropy, depend only on the probabilities the posterior assigns to the types, not on the parameter values of those types. Our variance reduction index weights the types by their parameter values, so it reflects how close the seller gets to a consumer’s parameter value rather than how well it identifies which type the consumer is.

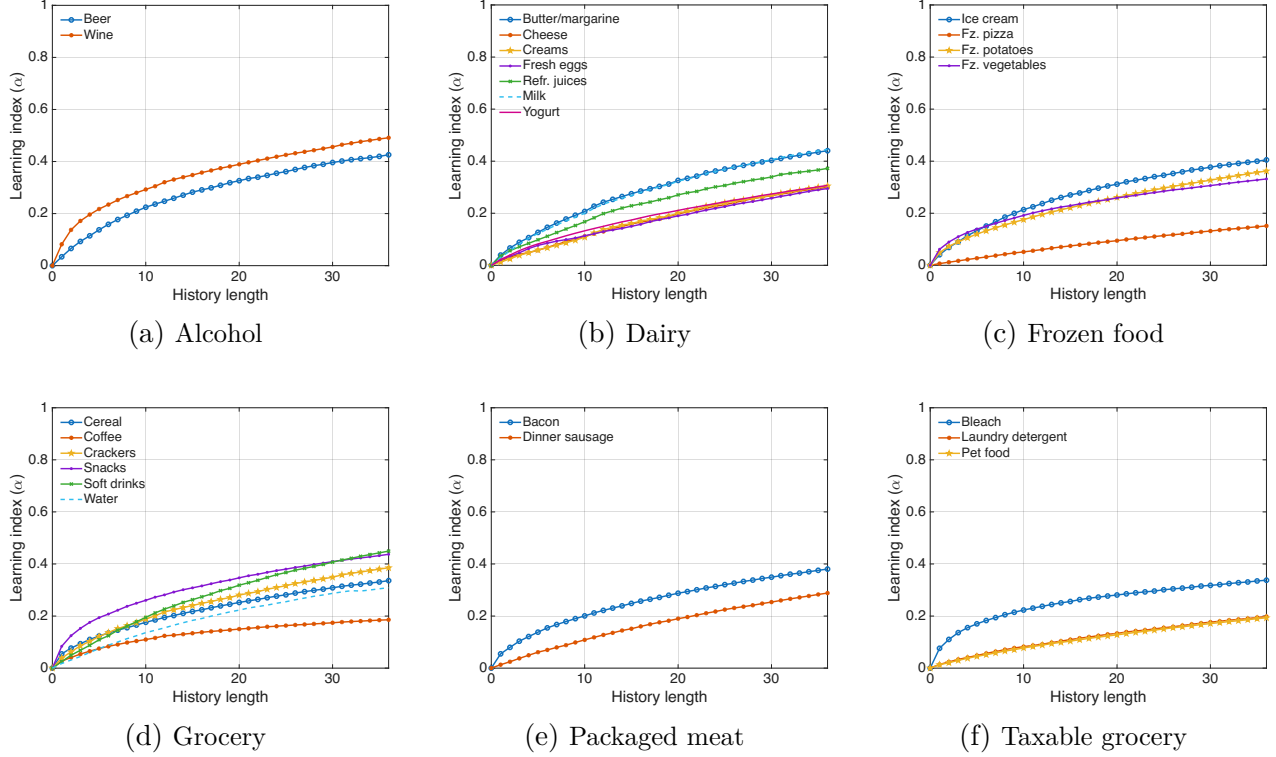
average posterior variance and the variance of the posterior mean across histories. Therefore, the index equals the share of the variance of η across consumers that the posterior means account for, the R^2 of a prediction of η from the purchase history.

Figure 4 presents the learning index for price sensitivity α across all product categories, grouped by department. The learning indices for the shopping occasion probability ρ and the category taste β are displayed in Figures C.1 and C.2. The learning indices increase monotonically with history length across all categories and parameters, but the pace of learning differs. Learning about ρ is very fast. Variance falls by 47% after one month and 65% after two, and the index reaches 0.998 after 36 months, indicating nearly complete learning. Even if ρ is not directly exploitable for personalized pricing (because it does not affect the optimal prices once the type is known), it provides information about preferences through the correlation with α and β .²⁴

The indices for α and β also increase monotonically with history length in every category, but learning remains incomplete: a 36-month window reduces α 's variance by 15–49% and β 's variance by 39–79%. The category taste is learned markedly faster than the price sensitivity: the average index at 36 months is 0.55 for β against 0.34 for α , and β is ahead in 22 of the 24 categories. The pace also varies widely across categories: half of the 36-month variance reduction in α is achieved within 4.9 months in bleach but requires 15.3 months in frozen pizza. Learning curves are concave in every category, with diminishing gains from each additional month of observation. Finally, the average index after one year is 0.19 for α and 0.39 for β . The first year delivers 53.8% and 69.8% of the 36-month reduction, respectively.

²⁴We quantify the value of indirect learning in Section 5.4.

Figure 4: Learning index for price sensitivity (α)



Notes: History length in months. Each figure shows one product department, with separate lines for the categories within it.

5.3 Dispersion of personalized prices

We now consider the distribution of personalized prices. For each consumer, we compute the unweighted average of personalized prices across the products in the category and express it as a percent deviation from its value under uniform pricing. [Table 2](#) reports its distribution in every category for 12-month histories. Personalized pricing lowers prices for most consumers. In every category, at least 85% of consumers pay less than under uniform pricing, and the median consumer receives a moderate discount of between 0.8% (bleach) and 8.0% (fresh eggs). Surcharges fall on a small group of consumers but are much larger. The maximum surcharge exceeds 20% in 22 categories and 50% in seven. The largest surcharges are much smaller at the level of the grocery basket, as surcharges in one category are partly offset by discounts in others: the largest basket surcharge is 24.2%, against up to 168% within a category. Across all categories, 92% of consumers pay less than under uniform pricing for

their basket.

Figure 5 shows how this distribution changes as information accumulates, from 3 to 12 and 36 months of histories, in two categories that differ in how fast discounts converge, beer and soft drinks.²⁵ Each panel ranks consumers from the highest- to the lowest-paying and plots their average personalized price, normalized by the uniform price.²⁶ The panel also reports

Table 2: Surcharges and discounts under personalized pricing, by category

Category	Surcharge/discount over uniform price (%)							Share below uniform
	Mean	Std. dev.	Min.	25th perc.	Median	75th perc.	Max.	
Alcohol								
Beer	-2.03	1.48	-4.13	-2.77	-2.35	-1.79	11.38	0.96
Wine	-5.91	4.60	-9.86	-7.06	-6.22	-5.52	168.44	0.98
Dairy								
Butter/margarine	-2.84	2.62	-5.26	-4.14	-3.50	-2.64	24.69	0.89
Cheese	-5.00	4.74	-10.75	-7.84	-5.76	-3.04	30.20	0.90
Creams	-4.11	2.18	-6.21	-5.17	-4.12	-3.44	96.55	0.97
Fresh eggs	-6.46	5.90	-11.53	-9.16	-7.96	-5.36	90.28	0.92
Refr. juices	-5.83	4.55	-10.78	-7.98	-6.52	-5.22	37.93	0.93
Milk	-3.52	3.29	-8.09	-5.22	-4.45	-3.38	23.49	0.92
Yogurt	-2.55	3.81	-8.13	-5.83	-2.07	-0.02	53.16	0.93
Frozen food								
Ice cream	-3.84	3.36	-5.49	-4.94	-4.69	-4.10	61.59	0.95
Fz. pizza	-1.82	1.07	-3.02	-2.44	-1.63	-1.58	23.45	0.97
Fz. potatoes	-2.83	1.96	-4.82	-4.00	-2.76	-2.25	32.69	0.94
Fz. vegetables	-6.09	4.40	-9.60	-8.45	-7.80	-5.85	37.47	0.93
Grocery								
Cereal	-4.02	2.77	-6.14	-5.39	-5.13	-3.98	27.01	0.93
Coffee	-1.93	1.41	-2.58	-2.26	-2.23	-1.97	32.25	0.97
Crackers	-3.86	4.22	-8.69	-5.85	-5.20	-3.79	30.64	0.90
Snacks	-4.06	5.43	-9.49	-6.97	-5.46	-2.78	27.96	0.88
Soft drinks	-4.81	6.96	-15.38	-8.44	-5.63	-3.28	42.54	0.85
Water	-3.41	3.22	-5.95	-4.85	-4.57	-3.76	29.62	0.91
Packaged meat								
Bacon	-4.62	3.60	-8.03	-6.39	-5.78	-4.62	51.74	0.93
Dinner sausage	-4.03	3.14	-6.49	-5.49	-5.30	-4.22	20.38	0.92
Taxable grocery								
Bleach	-0.78	0.44	-1.09	-1.00	-0.80	-0.73	18.62	0.98
Laundry detergent	-2.11	2.00	-3.59	-2.76	-2.50	-2.03	39.26	0.95
Pet food	-4.01	4.11	-9.46	-5.31	-4.21	-3.72	155.88	0.98
All categories (pooled)	-4.40	3.20	-9.41	-6.31	-5.34	-3.75	24.16	0.92

Notes: 12-month purchase histories. Distribution of unweighted average of personalized prices across the category's products, in percent deviation from its value under uniform pricing. Category rows weight each history by its unconditional probability. The last row is computed over the 20,748 cardholders: for each consumer, it reports the percent change in the price of a fixed basket containing every product of every category, in quantities equal to the products' market shares under uniform pricing. The last column reports the share of consumers paying less than under uniform pricing.

²⁵Figure C.3 shows the 12-month panel for all categories.

²⁶Consumers are ranked separately under personalized pricing and under perfect discrimination, so the figure compares the two distributions of prices rather than the prices that each consumer pays under the two

the perfect discrimination prices, the uniform price and the average marginal cost.²⁷

In both categories, personalized prices exceed the uniform price for a minority of consumers (4% in beer and 15% in soft drinks). The seller surcharges the few consumers it identifies as price-insensitive and discounts the many others. Even the largest discounts leave prices well above marginal cost. As histories lengthen, the personalized curve moves toward the perfect discrimination curve, but not at the same pace on both sides. In beer, the surcharges on the top consumers are in place within three months, whereas discounts build up slowly and remain far above the perfect discrimination prices after three years. In soft drinks, discounts converge as fast as surcharges, and after three years the lowest personalized prices reach the lowest perfect discrimination price.

Table C.1 extends the convergence analysis to all categories. For each category and history length, it reports the share of the gap between the uniform price and the perfect discrimination prices that personalized prices have closed, separately on the surcharge segment and on the discount segment.²⁸ Surcharges converge faster than discounts in 18 of the 24 categories with 3-month histories and in 17 with 12-month and 36-month histories. The surcharges take shape within a few months, while discounts keep building up for years.

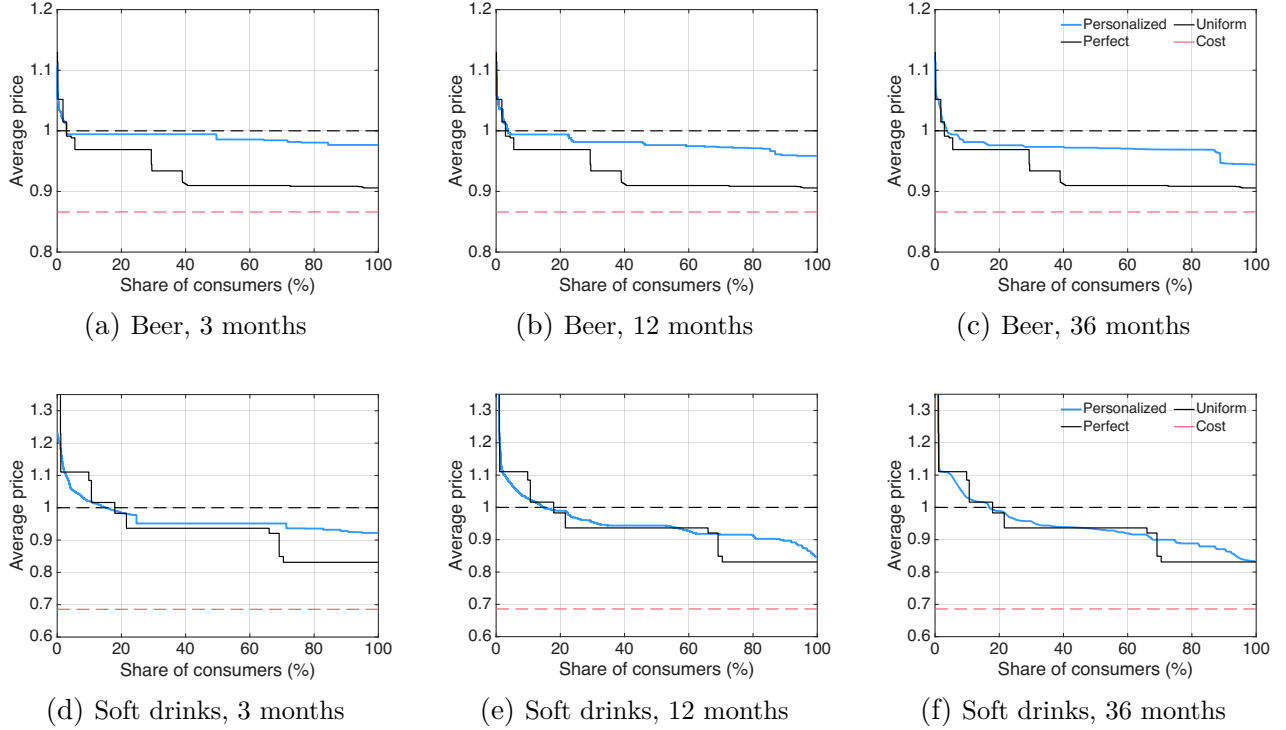
The asymmetry reflects the profit cost of pricing errors. Underpricing a price-insensitive consumer sacrifices margin on every purchase, while overpricing a price-sensitive consumer forgoes only the few purchases that this consumer would have made. Under uncertainty, the seller therefore keeps prices close to the uniform price by default and lowers them only once it is sufficiently confident that a consumer is price sensitive, whereas a moderate probability of facing a price-insensitive consumer already justifies a surcharge.

regimes.

²⁷The perfect discrimination curve is a step function with one step per taste pair (α, β) . The personalized price curves have one level per purchase history, and histories are numerous enough for the curves to appear continuous.

²⁸Consumers are ranked separately under personalized pricing and perfect discrimination. On the surcharge segment, the table pairs the highest personalized prices with the consumers that perfect discrimination surcharges, which maximizes the area we measure. The same holds on the discount segment. The shares are thus upper bounds on the shares computed consumer by consumer.

Figure 5: Distribution of prices, by history length (beer and soft drinks)



Notes: Average personalized and perfect price discrimination prices normalized by the average uniform price, across the consumer distribution; consumers ranked from highest-paying to lowest-paying. The dashed red line is the average marginal cost normalized by the average uniform price. In the soft drinks panels, the vertical axis is cut at 1.35; about 1% of consumers are priced above this level under perfect price discrimination and under personalized pricing with 12- and 36-month histories.

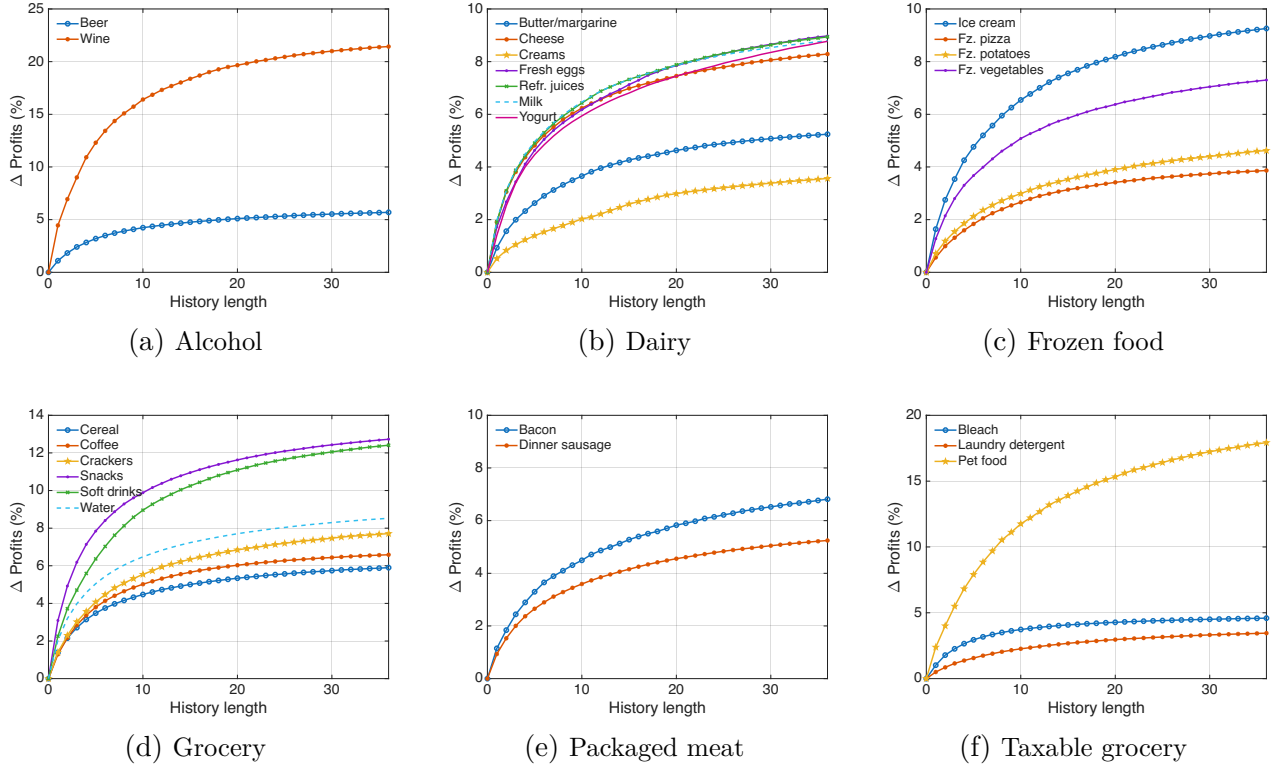
5.4 Profitability of personalized pricing

We study the profitability of personalized pricing as a function of the length of purchase histories. One year of purchase histories raises the seller's total profits by 7.1% over uniform pricing. As histories lengthen, personalized pricing approaches the perfect discrimination benchmark, and one-year histories already capture 65.9% of the perfect discrimination gain. [Figure 6](#) traces the gains in each category for histories of one month to three years.²⁹ The profitability of personalized pricing differs widely across categories. With 12-month histories, the gains range from 2.2% (creams) to 17.3% (wine).

The gains from personalization are concave in history length, with diminishing returns to each

²⁹[Table C.2](#) reports, by category, the profit gain under perfect price discrimination and the share of this gain captured with histories of 3, 12, and 36 months.

Figure 6: Impact of personalized pricing on profits (percent change from uniform)



Notes: Each line represents the evolution of the gains in profits from personalized pricing over uniform pricing, in percentage.

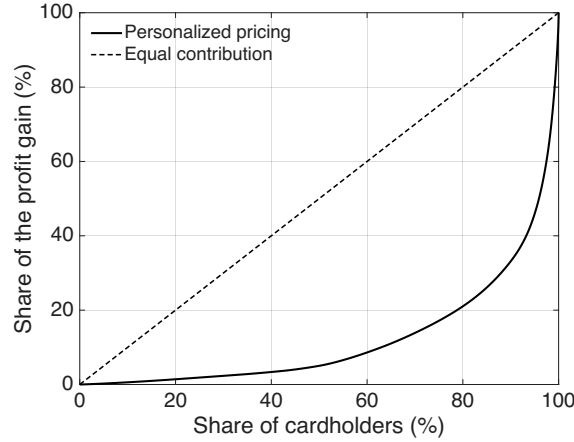
additional month of consumer data. The first three months deliver 41.5% of the three-year aggregate gain, and the first year delivers 77.4%. We find that profits run ahead of beliefs. After three years, the seller has resolved 33.9% of the variance of price sensitivities while capturing 85.2% of the perfect price discrimination gain. This difference highlights that not all purchase histories are equally valuable: the learning index weighs all consumers equally, whereas profits depend on identifying the small group of price-insensitive consumers to surcharge.³⁰ The concavity of the gains aligns with the previous section, where the seller surcharges price-insensitive consumers within a few months and discounts price-sensitive consumers more slowly.

We quantify how the seller's gains from price personalization are distributed across consumers.

³⁰We split the gain from personalized pricing by whether a history's average personalized price exceeds the uniform price. Histories with a surcharge carry more than 60% of the gain in every category and at every history length.

Figure 7 plots the Lorenz curve of the seller’s gains from 12-month histories across the 20,748 cardholders. The 5 percent most profitable cardholders account for 54.0% of the gains. The bottom 50 percent of cardholders account for only 5.0% of the gains. Table C.3 relates the gains to purchase frequency by category. In every category, most of the gains accrue from consumers with four or more purchases over the 12 months.

Figure 7: Concentration of the profit gains from personalized pricing across cardholders



Notes: Cumulative share of the seller’s profit gain from personalized pricing against the cumulative share of the 20,748 cardholders, ranked by the seller’s profit gain, for 12-month histories. The dashed line is the equal-contribution benchmark. The Gini coefficient is 0.755.

Profitability under incomplete histories. We now study the profitability of personalized pricing under alternative information structures to understand which part of the purchase history is most valuable. Table 3 reports the aggregate profit gains under four coarser information structures. First, we consider a seller that observes realized purchases but not store visits (“Transactions”), so that the absence of a purchase cannot be distinguished from the absence of a store visit. This information structure preserves 90% of the gains of the baseline seller (which observes complete histories) with 12-month histories, and 88% with 36-month histories. Second, we consider a seller that observes store visits and whether a purchase is made, but not which product is chosen (“No product identity”). This preserves 90% of the gains with 12- and 36-month histories, and more than half in every category. Third, we consider a seller that observes whether a purchase is made, but neither the products chosen nor the store visits (“Purchase incidence”). This preserves 80% of the gains with 12-month histories. Finally, we consider a seller that observes store visits only. Visits are informative

about preferences only through the correlation between the shopping occasion probability and the preference parameters, and store visits alone deliver 13% of the gains with 12-month histories. What matters most for the gains from personalization is the purchase frequency in the category. The products chosen and the store visits each have a limited impact on profitability.

Table 3: Profit gains from personalized pricing by information structure

Information structure	3 months		12 months		36 months	
	$\Delta\pi$ (%)	Share	$\Delta\pi$ (%)	Share	$\Delta\pi$ (%)	Share
Complete history	3.82	1.00	7.14	1.00	9.22	1.00
Transactions	3.74	0.98 [0.91, 0.99]	6.45	0.90 [0.82, 0.96]	8.10	0.88 [0.76, 0.95]
No product identity	3.49	0.91 [0.59, 1.00]	6.42	0.90 [0.54, 0.99]	8.30	0.90 [0.55, 0.98]
Purchase incidence	3.41	0.89 [0.54, 0.98]	5.68	0.80 [0.43, 0.94]	7.06	0.77 [0.41, 0.89]
Store visits	0.66	0.17 [0.01, 0.24]	0.95	0.13 [0.01, 0.20]	1.04	0.11 [0.01, 0.17]

Notes: Percent change in profits from uniform to personalized pricing depending on the information structure: the full purchase history (“Complete history”), purchases but not store visits (“Transactions”), store visits and whether a purchase is made but not which product (“No product identity”), whether a purchase is made but neither the product nor the store visits (“Purchase incidence”), or store visits alone (“Store visits”). “Share” is the gain under each information structure as a fraction of the gain under the complete history. The main figures aggregate the 24 categories, and brackets report the minimum and maximum across categories.

Price discrimination while learning. In our baseline, the seller holds the uniform price while it accumulates data and personalizes prices only afterwards. As a robustness check, we let the seller reprice every month in one randomly selected category, soft drinks. The seller charges each history the price implied by its current beliefs, so that consumers respond to personalized prices during the learning phase, and the seller learns from these responses.

We find that personalizing prices while learning does not reduce the gains. With 12-month histories, a seller that reprices every month earns 10.4% more than the uniform pricing benchmark, against 9.6% for the baseline seller. [Figure C.5](#) shows that the same holds at every history length up to three years. The exercise keeps the seller myopic, maximizing current profit given current beliefs without experimenting to accelerate learning. We examine the profitability of experimentation in [Section 5.6](#).

Learning across categories. The seller learns category by category, although a consumer’s types may be correlated across categories. Pooling histories across all categories is not feasible. The seller’s prior would be a joint distribution over the types of 24 categories, and its beliefs would depend on the purchase histories in all categories at once, which no recursion can track. We therefore analyze cross-category learning for two pairs of frequently co-purchased categories, milk and cereal, and snacks and soft drinks. We estimate the joint distribution of types in each pair, keeping fixed the marginal distributions in both categories (see estimation details in Appendix C.2.6). Types are positively correlated across categories, with coefficients of about 0.5 for category tastes and of 0.24 and 0.43 for price sensitivities. A seller pricing each category on the pooled histories would nevertheless gain little. Pooling raises the profit gains by 0.5 to 0.8 percentage points with 3-month histories and by 0.3 to 0.5 percentage points with 12-month histories (see Table C.4).

Learning from a misspecified prior. The seller knows the true distribution of consumer types. We relax this assumption and let the seller update from a misspecified prior that puts a weight of γ on a uniform distribution over the type grid and a weight of $1 - \gamma$ on the true population distribution. The seller sets personalized prices on the resulting posteriors and earns the profits implied by the true posteriors. Table 4 shows that purchase histories correct a misspecified prior over time. Without consumer data, the uniform price implied by the misspecified prior costs the seller 4.1% to 11.0% of profits, depending on γ . With 3-month histories, most of this loss is recovered, although the seller retains at most half of the gain from personalized pricing obtained under the correct prior. With 12-month histories it retains 73.0% to 89.4% of that gain, and with 36-month histories 92.3% to 96.8%. Purchase histories correct the prior because the prior enters the posterior as a fixed weight on each type, whereas the likelihood of the history increasingly favors the consumer’s true type as the history lengthens (see Appendix C.2.7 for the derivation).

Table 4: Profit gains from personalized pricing under a misspecified prior

	No history	3 months		12 months		36 months	
Seller’s prior	$\Delta\pi$	$\Delta\pi$	Share	$\Delta\pi$	Share	$\Delta\pi$	Share
Correct prior	0.00	3.82	100.0	7.14	100.0	9.22	100.0
Misspecified prior							
$\gamma = 0.25$	-4.14 [-30.30, -0.43]	1.86	48.7 [-525.8, 93.0]	6.38	89.4 [18.6, 97.2]	8.93	96.8 [80.1, 98.7]
$\gamma = 0.50$	-7.67 [-38.56, -1.31]	0.23	6.1 [-669.5, 79.9]	5.76	80.7 [-0.9, 93.4]	8.71	94.5 [74.6, 97.3]
$\gamma = 0.75$	-11.04 [-83.98, -2.30]	-1.18	-30.7 [-761.3, 64.8]	5.21	73.0 [-10.4, 88.9]	8.51	92.3 [72.1, 95.8]

Notes: All entries are in percent. $\Delta\pi$ is the change in profits relative to uniform pricing under the correct prior, when the seller updates from the correct prior or from a prior with weight γ on the uniform distribution over the type grid. “No history” is the change when the seller charges the uniform price implied by its prior. “Share” is the gain under the misspecified prior relative to the gain under the correct prior. The main figures aggregate the 24 categories, and brackets report the minimum and maximum of the share across categories.

5.5 Impacts of personalized pricing on consumer surplus

Aggregate consumer surplus falls at every history length, and the loss grows as histories lengthen. The loss is nevertheless small: 0.7% of consumer expenditure with 12-month histories and 1.2% under perfect price discrimination.³¹ Personalized pricing both transfers surplus from consumers to the seller and raises total welfare. With 12-month histories, consumers lose 39 cents for every dollar of profit that personalized pricing adds, so that total welfare rises by 61 cents per dollar of profit. The aggregate pattern is not driven by a few large categories. Consumer surplus falls at every history length in 21 of the 24 categories (Figure C.4). Total welfare rises in every category with histories of 12 months or longer.

These aggregate consumer surplus losses occur despite widespread discounts and overall market expansion (quantity sold expands by 3.5% to 11.4% across categories). This is the outcome Rhodes and Zhou (2024) derive when the uniform price excludes most consumers (low market coverage). Personalized pricing then expands demand and raises total welfare, but the consumers it brings into the market have low valuations and gain little, and those who already bought pay more, so that consumers lose in aggregate. Fewer than 15% of cardholders

³¹The level of consumer surplus depends on the normalization of the utility of the outside option. We therefore express changes in consumer surplus relative to consumer expenditure under uniform pricing.

buy in a category in a given month, so market coverage is low in our setting.

Table 5 reports the distribution of these surplus changes across consumers with 12-month histories. Consumer surplus falls on average, yet most consumers gain in every category.³² This is because gains are small and losses large. The median consumer gains between 0.2% and 1.5% of average expenditure, depending on the category, while the consumer at the first percentile loses between 14.6% and 140.1%. This asymmetry is the individual-level counterpart of the asymmetry of personalized prices in Table 2, where the surcharges on the consumers identified as price-insensitive are much larger than the discounts to everyone else. Whether these consumers can be compensated, and at what cost, is the question of Section 6.

³²The share of consumers who gain is close to, but distinct from, the share of consumers who pay less than under uniform pricing in Table 2. The latter compares the average of a consumer’s personalized prices across the category’s products with the uniform average, whereas surplus depends on the prices of the products the consumer likes.

Table 5: Distribution of consumer surplus changes at the individual level

Category	Δ Consumer surplus (% of avg. expenditure)						
	Mean	Share gaining	1st pct.	10th pct.	Median	90th pct.	99th pct.
Alcohol							
Beer	0.06	0.96	-32.51	0.13	0.51	4.21	14.56
Wine	-4.24	0.98	-21.96	0.95	1.13	9.84	44.31
Dairy							
Butter/margarine	-0.35	0.89	-54.21	-0.40	0.63	6.94	15.71
Cheese	-0.34	0.89	-111.13	-0.23	1.38	7.64	16.00
Creams	0.16	0.97	-35.93	0.60	0.65	5.04	10.83
Fresh eggs	-0.87	0.92	-116.88	0.88	1.52	11.01	16.77
Refr. juices	-0.34	0.93	-140.08	0.90	1.11	11.11	21.82
Milk	-0.51	0.90	-92.90	0.06	1.24	7.78	15.74
Yogurt	-0.91	0.74	-56.43	-0.02	0.29	10.49	21.97
Frozen food							
Ice cream	-1.28	0.94	-114.72	0.47	0.60	7.82	15.91
Fz. pizza	-0.14	0.97	-40.50	0.23	0.26	5.13	13.62
Fz. potatoes	-0.16	0.94	-64.80	0.33	0.54	7.89	17.05
Fz. vegetables	-0.59	0.93	-80.23	1.13	1.23	11.12	18.77
Grocery							
Cereal	-0.10	0.93	-62.97	0.61	0.81	6.35	8.45
Coffee	-0.42	0.96	-62.73	0.51	0.57	4.17	13.33
Crackers	-0.59	0.90	-77.26	0.01	0.82	6.36	13.32
Snacks	-1.16	0.88	-121.55	-1.61	1.16	7.01	13.27
Soft drinks	-1.26	0.86	-98.82	-6.47	1.38	8.42	15.81
Water	-0.46	0.90	-83.66	0.01	1.02	8.14	12.84
Packaged meat							
Bacon	-0.41	0.92	-107.12	0.74	0.97	8.98	12.70
Dinner sausage	-0.22	0.92	-102.54	0.70	0.93	9.12	18.18
Taxable grocery							
Bleach	-0.13	0.98	-14.64	0.18	0.23	3.29	10.45
Laundry detergent	-0.10	0.96	-28.20	0.17	0.30	3.64	12.30
Pet food	-3.04	0.98	-124.75	0.32	0.55	5.29	47.34

Notes: Change in consumer surplus at the individual level from uniform to personalized pricing (12-month histories), expressed as a percentage of the average consumer expenditure under uniform pricing in each category. Statistics are computed over belief nodes (purchase histories), weighted by their model-implied probabilities. Positive values indicate gains in surplus, negative values indicate losses; “Share gaining” is the share of consumers whose surplus rises. Because the change is normalized by average expenditure rather than by the consumer’s own expenditure, individual losses can exceed 100%.

5.6 Incentives to influence learning

Throughout our analysis, we maintain two assumptions about the behavior of the seller and consumers during the learning phase. First, the seller prices myopically, maximizing current expected profits given the unconditional distribution of types in each month. Second, consumers do not anticipate that their purchases will determine their future prices. The information accumulates as a by-product of consumers' choices. Either side could try to influence what the seller learns. The seller might distort prices to make purchase histories more informative, sacrificing profits today to personalize prices more effectively tomorrow. We call this strategy price experimentation. A consumer might likewise withhold purchases to appear price sensitive and obtain lower prices. We examine each incentive in turn.³³

Price experimentation by the seller. We consider the following two-period model. In the first period, the seller charges all consumers the same price vector $\mathbf{p}_1$, whose entry p_{j1} is the price of product j . The seller observes the one-period purchase histories generated at $\mathbf{p}_1$, and in the second period charges each consumer the personalized price implied by the posterior on that consumer. Let $\pi_1(\mathbf{p}_1)$ denote first-period profit and $\pi_2(\mathbf{p}_1)$ the expected second-period personalized profit. Second-period profit depends on $\mathbf{p}_1$ only, since the posteriors are determined by the histories generated at $\mathbf{p}_1$ and second-period prices are the optimal personalized prices given these posteriors. The benchmark is the myopic seller, whose first-period price is the uniform price $\mathbf{p}^U$.

We first compute the most profitable price experiment, ignoring its cost. The seller chooses $\mathbf{p}_1$ to maximize $\pi_2(\mathbf{p}_1)$ alone. The solution $\mathbf{p}_1^*$ is the best experiment, and the value of experimentation is $\pi_2(\mathbf{p}_1^*) - \pi_2(\mathbf{p}^U)$. The cost of generating more informative histories is $\pi^U - \pi_1(\mathbf{p}_1^*)$, the first-period profit forgone relative to uniform pricing. We report both in percent of the uniform profit π^U , and combine them into a break-even horizon, the number of months over which the gain would have to persist to repay the cost.

We then let the seller weigh the gain against the cost. A forward-looking seller chooses $\mathbf{p}_1$ to maximize $\pi_1(\mathbf{p}_1) + \pi_2(\mathbf{p}_1)$, where we assume a discount factor of 1. We compare its profits over the two periods with those of the myopic seller, $\pi^U + \pi_2(\mathbf{p}^U)$. Appendix C.3 discusses the implementation.

³³We study specific strategies where the seller and consumers try to influence learning in isolation from each other. Solving for the full equilibrium, as in [Bonatti and Cisternas \(2020\)](#), is left for future work.

Table 6 shows that the gain from price experimentation is small relative to its cost. In aggregate, the best experiments raise second-period profits by 0.46% of uniform profit, at a cost of 52.0% of first-period profits. It takes 112 months to recoup the seller’s forgone profits, and 26 months in the most favorable category (pet food). These horizons are lower bounds, as they assume a unit discount factor and gains that persist undiminished. A forward-looking seller, which maximizes the profits of both periods, earns 0.001% more than the myopic seller, and at most 0.005% more in any category.³⁴

Table 6: The value of price experimentation

	Change in profit relative to the myopic seller			Break-even (months)
	Second period (%)	First period (%)	Two periods (%)	
Best experiment	0.46 [0.04, 1.49]	-52.0 [-94.1, -16.2]	-25.5 [-46.6, -8.0]	112 [26, 833]
Forward-looking seller	0.003 [0.000, 0.019]	-0.002 [-0.009, 0.000]	0.001 [0.000, 0.005]	

Notes: Changes are relative to the myopic seller, which charges the uniform price in the first period. The best experiment maximizes second-period profit, and the forward-looking seller maximizes the sum of first-period and second-period profits. Changes in first-period and second-period profit are in percent of uniform profit, and the change in two-period profit is in percent of the myopic seller’s two-period profit. “Break-even” is the number of months of personalized pricing needed to repay the cost of the best experiment, undiscounted. The main figures aggregate all categories, and brackets report the minimum and maximum across categories.

We run two robustness checks and find the conclusion unchanged. First, a longer experiment does not help. In fresh eggs, the only category with two products, a seller that experiments for two periods charges the same prices in both and gains 0.18% of the myopic seller’s profit. Second, experimentation remains costly once the seller holds some purchase histories. In four categories, experiments tailored to three-month histories break even faster than the initial experiment in three of them, but still take at least 29 months. Appendix C.3 gives the details.

Withholding purchases by consumers. We now turn to consumers. They could manipulate their purchase patterns in order to obtain lower personalized prices afterwards. We assume that they do not manipulate their shopping occasions, and that their strategy is to make no purchase in the category at any shopping occasion over 12 months. In all but

³⁴Table C.5 reports the results by category.

two categories (laundry detergent and pet food), this is the strategy that obtains the lowest personalized price.³⁵ The seller does not anticipate this behavior and sets its personalized prices as in our baseline, so the exercise is not an equilibrium analysis. The cost is the surplus that the consumer forgoes in the category over the 12 months. The gain is the increase in the consumer’s monthly surplus in the months that follow, relative to the personalized prices based on the true purchase history. As we did for the seller, we combine the two into a break-even horizon, the number of months of lower prices needed to recover the cost.

The gain from withholding all purchases is small relative to its cost. The consumers who would obtain a lower price by withholding would receive an extra discount between 0.8% (bleach) and 11.6% (soft drinks) on average. Recovering the forgone surplus takes a long time, even for the consumers that gain the most. In 23 categories, the shortest break-even horizon lies between 14.3 months (snacks) and 26.2 months (laundry detergent). A consumer who makes no purchase for 12 months therefore gains only if the low prices then last for more than 12 months. Soft drinks is the exception. The shortest horizon there is 9.3 months, and the types that break even within 12 months account for 6.9% of the consumers who purchase in the category.³⁶ The break-even horizons are lower bounds, as they assume a unit discount factor and lower prices that persist undiminished. Once the consumer purchases again, the seller’s beliefs move back toward the consumer’s true type. Appendix C.4 formalizes the exercise and reports the complete results.

6 Design of Loyalty Programs

Our analysis so far has endowed the seller with every consumer’s purchase history. In practice, these data exist because consumers enroll in a loyalty program, and consumers might refuse to surrender histories that will be used to price against them. We therefore study the design of a voluntary loyalty program. Its design depends on what each consumer gains or loses from personalized pricing, and on what the seller gains from that consumer’s data.

Figure 8 compares, for each cardholder in our sample, the change in surplus with the change in the seller’s profit from personalization, in dollars per month.³⁷ The surplus change is

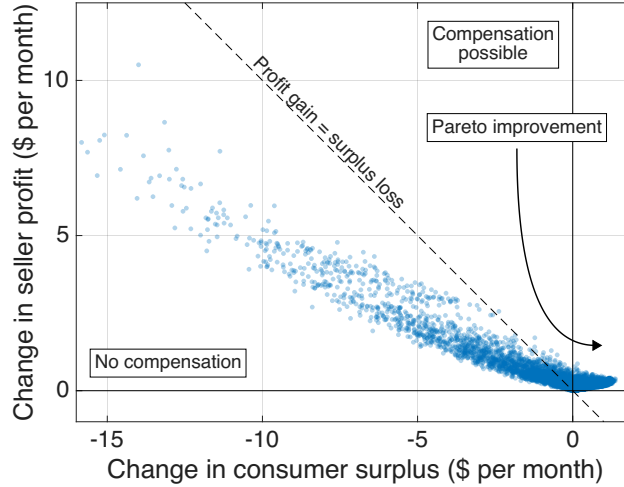
³⁵We do not evaluate strategies that forgo only some purchases. They cost less and could therefore break even sooner, although they obtain a smaller discount.

³⁶The share is computed among the consumers with at least one purchase in the category over the 12 months, since a consumer with no purchase has nothing to withhold.

³⁷Summing gains and losses across categories requires the joint distribution of a consumer’s histories, which

the consumer’s willingness-to-pay to share the data, and the profit change is the seller’s willingness-to-pay to acquire purchase histories. The figure partitions consumers into three groups. For 84.1% of consumers, both the consumer and the seller gain. Another 3.3% lose, but the seller’s gain is larger, so compensation is feasible. The remaining 12.5% lose more than the seller gains, so compensation is infeasible. The seller’s gains come mostly from this last group, which brings 67.1% of its profit gain.³⁸

Figure 8: Change in profits vs. change in consumer surplus



Notes: Each dot represents one consumer. *Pareto improvement*: both parties gain. *Compensation possible*: seller gains exceed consumer surplus losses. *No compensation*: consumer surplus losses exceed seller gains. Results based on 12-month purchase histories, summed over the 24 categories. The axes are truncated at a surplus loss of \$16 and a seller gain of \$12.5 per month. Twelve of the 20,748 consumers fall outside, losing up to \$39.0 or bringing the seller up to \$26.9, and are not shown but are included in every statistic reported in the text.

In what follows, we let consumers choose whether to enroll in a loyalty program that shares their data, which enables the seller to track their purchases and personalize prices. The seller offers a monthly uniform transfer b to every consumer who enrolls.³⁹ Enrollees reveal

we do not model. We therefore evaluate each cardholder at the 12-month histories observed in the data, which preserve the correlation of purchases across categories.

³⁸The partition classifies consumers by their expected consumer surplus changes, $\sum_{k=1}^{24} \sum_{d=1}^{D_k} \phi_{d|\mathbf{h}_i^k}^k [CS_d^k(\mathbf{p}^k(\mathbf{h}_i^k)) - CS_d^k(\mathbf{p}^{U,k})]$. We could instead classify each combination of types $(d_1, \dots, d_{24})$ by the sign of $\sum_{k=1}^{24} [CS_{d_k}^k(\mathbf{p}^k(\mathbf{h}_i^k)) - CS_{d_k}^k(\mathbf{p}^{U,k})]$ and weight the combinations by their posterior probabilities. We compute this with draws from the joint posterior and obtain a very similar partition: 83.6% gain, 3.7% lose less than the seller gains, and 12.6% lose more. The last group brings 66.0% of the seller’s gain.

³⁹Monthly and annual fees are equivalent for enrollment decisions when the environment is stationary, that is, when costs and product qualities are identical in every month.

their 12-month purchase histories and face personalized prices in the following period, while non-enrollees keep their data private and face uniform prices. A positive transfer is a reward for membership, and a negative transfer is a membership fee. We express it in percent of the average consumer expenditure over our 24 categories (\$10.1 per month). Consumers know their own preferences and anticipate the personalized prices implied by their purchase histories. A consumer enrolls if the surplus at personalized prices plus the transfer exceeds the surplus at the uniform prices set for non-enrollees, both summed over all categories. We do not observe preferences, so we compute each consumer’s probability of enrolling under the posterior distribution of types given purchase histories (Appendix C.5).

The seller faces the following trade-off. On the one hand, the consumers who bring the largest profit gains are those who lose from personalized pricing and must be paid to enroll. On the other hand, consumers who gain from personalized pricing would pay to enroll, and the seller can extract part of their surplus. Because the transfer is the same for everyone, the seller must trade off attracting the first group against extracting surplus from the second.

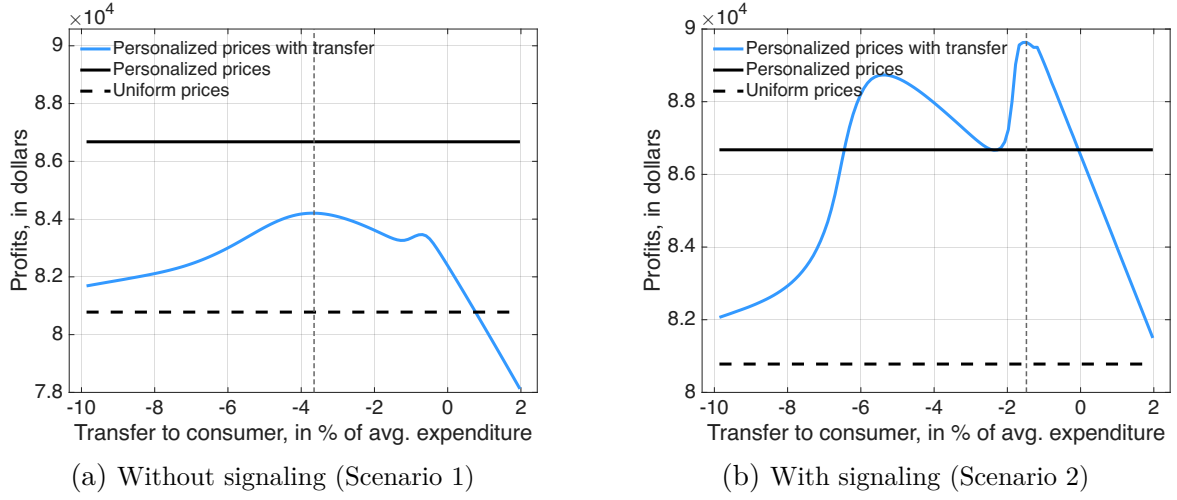
We assume that the seller does not update its beliefs upon enrollment and learns only from purchase histories, category by category.⁴⁰ However, we consider two scenarios that depend on whether the seller updates its beliefs upon non-enrollment. In the first scenario, the seller does not update its beliefs upon non-enrollment, and non-enrollees keep facing the baseline uniform prices, as in Conitzer et al. (2012). This is realistic when non-enrollees include many new consumers, or consumers who do not make an active enrollment decision. In the second scenario, the seller learns from non-enrollment, since consumers who withhold their data signal that they would lose from enrolling, as in Ali et al. (2023). Consumers hold rational expectations about this price, and enrollment decisions and the non-enrollee price are solved jointly as a fixed point (see Appendix C.5).⁴¹

Scenario 1 Panel (a) of Figure 9 shows the seller’s total profits under Scenario 1 as a function of the transfer. Profits are maximized at a fee of 3.65% of average expenditure. At this fee, 25.8% of consumers enroll and profits exceed uniform-pricing profits by 4.2%. The

⁴⁰Updating beliefs upon enrollment would make the personalized prices depend on the transfer, since the set of enrollees does. Every history would then have to be repriced at every transfer of the grid search, whereas the personalized prices are computed once.

⁴¹At each transfer, we compute the equilibrium reached by iterating from no enrollment. Starting from other enrollment sets reaches the same equilibrium except at a few fees slightly larger than the optimal fee (see Appendix C.5). Our algorithm also rules out the equilibria in which prohibitive prices that no consumer faces induce full enrollment.

Figure 9: Loyalty program profits by transfer



Notes: Total seller profits over all categories and consumers. The transfer is paid monthly and expressed in percent of average consumer expenditure (\$10.1 per month). The vertical dashed lines indicate the optimal transfers (-3.65% in Scenario 1, -1.5% in Scenario 2). Profits under uniform and full personalized pricing are shown as benchmarks.

fees account for 70.5% of the gain. The rest comes from price discrimination on enrollees, who are the consumers receiving discounts under personalized pricing. Profits remain 2.9% below those obtained when every consumer receives personalized prices. This is because the fee keeps out the price-insensitive consumers, who are the most profitable.

Scenario 2 Panel (b) of Figure 9 shows the results with signaling. The optimal fee is lower, 1.5% of average expenditure. At this fee, 94.5% of consumers enroll. Profits rise above uniform-pricing profits by 11.0% and above those obtained when every consumer receives personalized prices by 3.4%. Fees account for 40.5% of the gain and price discrimination on enrollees for 53.3%. The repricing of non-enrollees accounts for the remaining 6.2%.

Consumers who opt out reveal themselves to be price-insensitive, so the seller charges non-enrollees higher uniform prices, which makes enrollment more attractive. As more consumers opt in, the non-enrollee pool becomes less price sensitive and its uniform price rises, inducing further enrollment. Enrollment rises steeply as the fee falls toward its optimal level, from 55% at a fee of 2.1% to 71% at 1.9% and 94.5% at the optimal fee (Figure C.6). The consumers who stay out are those who lose the most from personalized pricing: they would lose \$1.28 per month on average at personalized prices, against \$0.04 for enrollees, and the seller sets their

prices 18.5% above the uniform prices on average. The program leaves 66.0% of consumers worse off than under uniform pricing.

Whether voluntary enrollment protects consumers from personalized pricing depends on what the seller learns from non-enrollment. When non-enrollment is uninformative, the fee is paid by the price-sensitive consumers who obtain discounts, and the others stay out and are unaffected. When non-enrollment reveals price insensitivity, most consumers pay the fee and two thirds end up worse off than under uniform pricing.

7 Conclusion

We propose a framework to study the value of consumer purchase data in retail markets, combining Bayesian learning about consumer preferences from purchase histories with behavior-based personalized pricing. We estimate the framework’s market primitives using data from a U.S. supermarket covering 24 product categories that represent about 20% of the store’s sales, leveraging the purchase histories that the store’s loyalty program records for 20,748 consumers.

We find that purchase histories are informative. After one year, the firm resolves on average 18.6% of the variance of price sensitivities and 38.7% of the variance of category tastes. Personalized pricing increases total profits across all categories by 7.1% with one year of data. Total welfare rises slightly, but aggregate consumer surplus falls by 0.7% of consumer expenditure. This small aggregate loss masks important distributional effects. While a large majority of price-sensitive consumers experience small gains, personalized pricing imposes substantial losses on a small group of consumers whom the seller identifies as price-insensitive. These concentrated losses raise questions for consumer data protection policy.

We apply our framework to loyalty program design. The optimal program charges a membership fee, which extracts part of the surplus that lower personalized prices give to price-sensitive consumers. When the firm treats non-enrollment as a signal of price insensitivity and raises the prices of non-enrollees, most consumers enroll and profits rise further. Two thirds of consumers then end up worse off than under uniform pricing. Voluntary enrollment therefore protects consumers only when the firm does not learn from opting out.

Two important limitations point to directions for future research. First, our analysis abstracts from competition: understanding how firms compete with asymmetric consumer data would

provide a more complete picture of the welfare effects of behavior-based pricing. Second, while we show that the seller has little incentive to experiment with prices to accelerate learning and consumers little incentive to withhold all their purchases, we evaluate each deviation on its own, holding the behavior of the other side fixed. The equilibrium in which both the seller and consumers act strategically remains unexplored.

References

- Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 54(2):442–492.
- Acquisti, A. and Varian, H. R. (2005). Conditioning prices on purchase history. *Marketing Science*, 24(3).
- Adams, B. and Williams, K. R. (2019). Zone pricing in retail oligopoly. *American Economic Journal: Microeconomics*, 11(1):124–156.
- Ali, S. N., Lewis, G., and Vasserman, S. (2023). Voluntary disclosure and personalized pricing. *The Review of Economic Studies*.
- Anderson, E. T. and Simester, D. I. (2010). Price stickiness and customer antagonism. *The Quarterly Journal of Economics*, 125(2):729–765.
- Asker, J., Fershtman, C., and Pakes, A. (2024). The impact of artificial intelligence design on pricing. *Journal of Economics & Management Strategy*, 33(2):276–304.
- Atalay, E., Frost, E., Sorensen, A., Sullivan, C., and Zhu, W. (2026). Scalable demand and markups. *Journal of Political Economy*, 134(8):000–000.
- Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–957.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2nd edition.
- 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.
- Berto Villas-Boas, S. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *The Review of Economic Studies*, 74(2):625–652.
- Bhattacharya, V., Illanes, G., Kreps, A. A., Salas, J. D., and Stillerman, D. (2026). A large-scale evaluation of merger simulations.

- Bonatti, A. and Cisternas, G. (2020). Consumer scores and price discrimination. *The Review of Economic Studies*, 87(2):750–791.
- Bonnet, C. and Dubois, P. (2010). Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance. *The RAND Journal of Economics*, 41(1):139–164.
- Calvano, E., Calzolari, G., Denicolò, V., and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297.
- Conitzer, V., Taylor, C. R., and Wagman, L. (2012). Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science*, 31(2):277–292.
- 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. H., and Stiebale, J. (2025). Rising markups and the role of consumer preferences. *Journal of Political Economy*, 133(8):2462–2505.
- 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. (2025). Estimating complementarity with large choice sets: An application to mergers. *The RAND Journal of Economics*, 56(4):689–707.
- 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.
- Fox, J. T., Kim, K. i., Ryan, S. P., and Bajari, P. (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics*, 2(3):381–418.

- Fudenberg, D. and Tirole, J. (2000). Customer poaching and brand switching. *RAND Journal of Economics*, pages 634–657.
- Fudenberg, D. and Villas-Boas, J. M. (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436.
- Gandhi, A. and Houde, J.-F. (2026). Measuring substitution patterns in differentiated-products industries. *NBER Working paper*, (w26375).
- Grieco, P. L., Murry, C., Pinkse, J., and Sagl, S. (2025). Optimal estimation of discrete choice demand models with consumer and product data.
- Hart, O. D. and Tirole, J. (1988). Contract renegotiation and coasian dynamics. *Review of Economic Studies*, 55(4):509–540.
- Hitsch, G. J., Hortacsu, A., and Lin, X. (2021). Prices and promotions in us retail markets. *Quantitative Marketing and Economics*, 19(3):289–368.
- Ichihashi, S. (2020). Online privacy and information disclosure by consumers. *American Economic Review*, 110(2):569–595.
- Kehoe, P. J., Larsen, B. J., and Pastorino, E. (2020). Dynamic competition in the era of big data. Working paper, Stanford University.
- Matcham, W. (2025). Risk-based borrowing limits in credit card markets. *Available at SSRN 4926974*.
- Miravete, E. J. (2026). On anticompetitive third-degree price discrimination.
- Miravete, E. J., Seim, K., and Thurk, J. (2024). Targeted pricing and the shape of demand. Working paper, November 2024.
- Morrow, W. R. and Skerlos, S. J. (2011). Fixed-point approaches to computing Bertrand-Nash equilibrium prices under mixed-logit demand. *Operations Research*, 59(2):328–345.
- Rhodes, A. and Zhou, J. (2024). Personalized pricing and competition. *American Economic Review*, 114(7):2141–2170.
- 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.

- Sagl, S. (2023). Dispersion, discrimination, and the price of your pickup. Job market paper, Pennsylvania State University.
- Shiller, B. R. (2020). Approximating purchase propensities and reservation prices from broad consumer tracking. *International Economic Review*, 61(2):847–870.
- Shiller, B. R. (2025). Inconspicuous personalized pricing. *The Journal of Industrial Economics*, 73(4):608–619.
- Smith, A. N., Seiler, S., and Aggarwal, I. (2023). Optimal price targeting. *Marketing Science*, 42(3):476–499.
- 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.
- Villas-Boas, J. M. (1999). Dynamic competition with customer recognition. *The RAND Journal of Economics*, pages 604–631.
- Villas-Boas, J. M. (2004). Price cycles in markets with customer recognition. *The RAND Journal of Economics*, 35(3):486–501.
- Wald, A. (1950). *Statistical Decision Functions*. Wiley, New York.
- Waldfogel, J. (2015). First degree price discrimination goes to school. *The Journal of Industrial Economics*, 63(4):569–597.

Appendices

A Data

A.1 Construction of the sample

This appendix details the construction of the estimation sample summarized in Section 3: the selection of consumers and categories, the cleaning of the transaction data, the definition of products, and the construction of prices, wholesale prices, and margins.

Consumer selection. We limit our analysis to consumers with a loyalty card who made at least one purchase at our chosen store in 2018. We drop five consumers with unreasonably high purchase volumes (their annual spending exceeds \$15,600 per year, equivalent to \$300 per week). Our final sample contains 20,748 consumers.

Category selection. We focus on six of the 56 departments, namely alcohol, dairy products, frozen food, grocery, packaged meat, and taxable (non-food) grocery. We leave out the departments that sell unbranded or store-prepared goods, such as fresh meat, produce, and bakery, where products cannot be defined by manufacturer. Within the six departments, we select categories in two steps, using the sales of all 853 stores so that the selection does not depend on our chosen store. First, we keep the categories whose sales exceed 5% of department revenue (2.5% for the grocery and taxable grocery departments, which comprise many small categories). Second, we keep the categories whose products can be expressed in a common standard unit, such as fluid ounces, counts, pounds, or loads.

This screening leaves 26 candidate categories. Our final list of 24 categories further excludes spirits, which have no recorded sales in our chosen store, and baby formula, whose sales are concentrated among a small number of parents of infants.⁴²

Final sample. Within the 24 categories, we apply two restrictions to select the set of UPCs and transactions. First, we drop the UPCs that cannot be converted to the category’s standard unit, either because no standard package size is recorded for their subcategory (the finest level of the product classification below the category) or because their recorded unit of measure, such as “pack” or “each”, does not identify a quantity in the standard unit. These

⁴²Only 383 of the 20,748 sample consumers ever purchase baby formula, and 52 of them account for half of the category’s sales.

UPCs account for 1.6% of revenue in the selected categories. Second, we drop the transactions with non-positive sales or quantities, such as refunds and register corrections. They represent 0.7% of transaction lines and less than 0.1% of revenue.

Product definition. Within each category, we aggregate UPC-level data into products in two steps. First, we standardize all UPCs to common units (e.g., ounces for cereal, fluid ounces for milk). Second, we aggregate by manufacturer, treating each manufacturer as a single product in the choice set.⁴³ This keeps the choice set small and the counterfactual simulations tractable, and it matches the level at which retail prices and promotions are typically set. We consolidate all manufacturers with revenue shares below 5% into a single composite product.

The model allows at most one purchase per category and shopping occasion. In 19.0% of the consumer-category-months with a purchase, the consumer buys from several manufacturers, and we assign the choice to the manufacturer with the largest expenditure. Since the model focuses on discrete product choice rather than quantity decisions, every purchase is of the same quantity, the category’s average monthly quantity per consumer conditional on purchase, and we scale prices and wholesale prices proportionally.

Description of the final sample. Table A.2 provides summary statistics for the 24 product categories in our sample. The number of products per category (including the composite product) ranges from 2 to 8, with a mean of 4.7. Wholesale margins vary across categories, from 15.1% of the retail price (beer) to 33.1% (creams), with an average of 26%. Column 7 reports average monthly sales: soft drinks and snacks are the largest categories by value, bleach and creams the smallest.

Prices. The model requires a single price for each product and month. Products group several UPCs, and the mix of UPCs that consumers buy changes from month to month, so a simple average price would confound genuine price variations, such as promotional discounts, with changes in the UPC mix. To avoid this, we measure the price of a product over a fixed basket of UPCs: those purchased in every month of the year, weighted by their annual quantities.⁴⁴ Price variations capture promotional discounts and are, by construction,

⁴³For example, 74 Kellogg’s cereal UPCs become one “Kellogg’s cereal” product.

⁴⁴For two products, no UPC is sold in every month. In that case, the basket contains all the product’s UPCs, weighted by annual quantities, and the weights are renormalized each month over the UPCs sold that

unaffected by changes in the mix of UPCs purchased.

Wholesale prices. We observe the invoice cost and the quantity of every delivery of each UPC to the store in 2017–2018. We calculate the wholesale unit price of each delivery as the invoice cost divided by the quantity delivered. Then, to each sale of a UPC, we assign the wholesale unit price of the most recent delivery on or before the sale date. The wholesale price of the UPC in a month is the average of these unit prices over the month’s sales, weighted by units sold. In the rare months in which a UPC has no delivery before its sales, we use its median wholesale price over 2018.⁴⁵

Observed margins. The margin of a product in a month is the average of price minus wholesale price over its UPCs, using the same basket weights as the prices. Two kinds of UPC-months are excluded from the calculation of the product margin but used for the price. First, UPCs without any recorded delivery have no wholesale price.⁴⁶

Second, we exclude negative margins.⁴⁷ The wholesale price reported in [Table A.2](#) is the price minus the margin. The estimation uses the margins directly, and the wholesale prices serve to validate the recovered marginal costs.

Consumer spending at each stage. [Table A.1](#) provides summary statistics on consumers’ purchases at each stage of the sample construction. Loyalty program members (column 2) spend \$30.8 per transaction on average, close to the \$32.9 of the full sample (column 1), though the full sample has a heavier right tail. Members spend on average \$41.8 per month at the supermarket, and \$102.0 in the months where they visit the store. Restricting to the selected departments (column 3) and to the selected categories (column 4), average monthly spending drops to \$25.1 and \$11.0. The selected categories thus represent approximately one quarter of members’ spending. The final sample (column 5) keeps 98.4% of the revenue of the selected categories, and monthly spending is nearly unchanged at \$10.8. Column 6 reports the spending implied by the objects we use in the estimation, namely the purchase histories and our product prices. Implied spending is \$11.4 per month, close to the \$10.8 observed in column 5.

month, so the price of these two products can also vary with the mix of UPCs sold.

⁴⁵This imputation affects 6 of the 23,509 UPC-months in the baskets.

⁴⁶They account for less than 2% of basket quantity for every product except four.

⁴⁷Invoice costs are gross of vendor allowances and other rebates that we do not observe, so the measured margin is understated and, for some UPC-months, the price falls below the measured wholesale price.

Table A.1: Consumer spending distribution at each stage of sample construction

	Full dataset (1)	Selected consumers (2)	Selected departments (3)	Selected categories (4)	Final sample (5)	Purchase histories (6)
Consumer spending per transaction						
Mean	32.9	30.8	21.0	12.5	12.4	–
S.D.	61.0	38.2	26.5	13.6	13.4	–
Median	16.4	17.9	12.0	8.2	8.1	–
99th pct.	230.0	180.8	128.7	64.1	63.0	–
Observations	397,377	337,523	296,248	217,755	216,367	–
Consumer spending per month						
Mean	–	41.8	25.1	11.0	10.8	11.4
S.D.	–	98.2	62.6	27.9	27.5	22.0
Median	–	0.0	0.0	0.0	0.0	0.0
99th pct.	–	444.7	288.6	129.3	126.5	94.8
Observations	–	248,976	248,976	248,976	248,976	248,976
Consumer spending per month, conditional on a visit						
Mean	–	102.0	64.8	32.6	32.1	33.8
S.D.	–	131.9	86.9	40.2	39.5	26.0
Median	–	58.7	35.4	19.6	19.4	26.7
99th pct.	–	601.2	396.9	186.1	183.1	114.5
Observations	–	102,073	96,298	83,944	83,733	83,733
Total sales (million)	13.07	10.41	6.24	2.73	2.69	2.83
# of consumers	All	20,748	20,748	20,748	20,748	20,748

Notes: Spending in 2018 dollars. (1): full dataset. (2): loyalty program members. (3): six selected departments. (4): 24 selected categories. (5): final sample. (6): final sample, implied from final product prices and purchase histories.

Table A.2: Summary statistics by product category

Category	# Products	Standard unit	Monthly quantity (units)	Standardized price (\$)	Standardized wholesale price (\$)	Monthly sales (\$)
Alcohol						
Beer	4	144 oz	1.48	16.13	13.69	6,411
Wine	7	25.4 oz	3.14	18.02	13.50	5,480
Dairy						
Butter/margarine	5	16 oz	2.12	4.03	2.93	6,059
Cheese	4	8 oz	3.45	8.11	5.73	21,224
Creams	4	32 oz	1.41	5.10	3.41	2,700
Fresh eggs	2	12 ct	1.83	3.84	2.62	7,382
Refr. juices	4	59 oz	3.22	5.51	3.70	8,690
Milk	5	64 oz	2.56	5.43	3.74	14,682
Yogurt	4	5.3 oz	7.56	5.29	3.93	3,018
Frozen food						
Ice cream	5	48 oz	1.9	6.54	4.88	5,924
Fz. pizza	5	10 oz	4.81	7.74	6.11	3,388
Fz. potatoes	4	32 oz	1.67	4.56	3.24	3,453
Fz. vegetables	4	12 oz	4.63	6.11	4.12	8,781
Grocery						
Cereal	5	12 oz	2.64	6.57	5.33	8,203
Coffee	6	12 oz	2.07	7.60	6.11	5,027
Crackers	5	12.4 oz	2.1	5.29	4.12	8,517
Snacks	3	8 oz	3.55	8.27	6.39	24,324
Soft drinks	5	144 oz	3.21	11.56	8.25	38,277
Water	5	128 oz	5.32	6.76	4.84	11,818
Packaged meat						
Bacon	8	24 oz	1.28	10.42	7.45	11,443
Dinner sausage	7	14 oz	3.44	8.67	5.85	11,980
Taxable grocery						
Bleach	4	48 ld	2.24	3.43	2.62	1,524
Laundry detergent	4	32 ld	2.21	8.42	6.96	2,910
Pet food	4	16 lb	0.82	12.01	9.27	2,734

Notes: “Standardized price” and “standardized wholesale price” are the average price and wholesale price of one unit, standardized to the category’s standard unit (column 3) and scaled by the average monthly quantity. Both are averaged over products and months using fixed basket weights. “Monthly quantity” is the average number of standardized units a consumer purchases in a month, conditional on purchasing at least one unit. “ld” denotes laundry loads. “Monthly sales” is total category sales in dollars by the sample consumers, averaged over the twelve months of 2018.

A.2 Price uniformity

This appendix shows that the store’s prices are effectively uniform across consumers. We compare the prices that different consumers pay for the same UPC on the same day, which we call a UPC-day, over all the products sold at the store. We define the unit price as the expenditure on a transaction line divided by the quantity purchased, and we exclude UPCs sold by weight. Consumers pay the same price in 92.6% of the 416,104 UPC-days with at least two transactions. We call the remaining 7.4% multi-price UPC-days.

Within multi-price UPC-days, 70.6% of transactions occur at the modal price, the most frequent price within a UPC-day. Anonymous shoppers pay above the modal price in 70.4% of their transactions on these UPC-days, against 12.4% for cardholders. Within UPC-days where both groups purchase, cardholders pay on average \$0.28 less per unit than anonymous shoppers (column 1 of [Table A.3](#)). We take this as evidence that some promotional discounts are tied to loyalty program membership. Quantity discounts are a second source of price dispersion: unit prices fall with the quantity purchased, by about \$0.11 per additional unit on average (column 2). When we remove anonymous shoppers and multi-unit transactions, 96.3% of UPC-days have a single price, and 98.4% of the transactions occur at the modal price.

Finally, we test whether the store conditions prices on purchase histories. On the single-unit purchases of cardholders, we regress the unit price on the consumer’s number of shopping months and, separately, on the logarithm of the consumer’s annual expenditure at the store (columns 3 and 4). We include UPC-day fixed effects to control for promotional discounts that apply to all consumers on a UPC-day. An additional shopping month changes the unit price by a tenth of a cent, and a doubling of annual expenditure raises it by less than half a cent. The first coefficient is not statistically significant, and the second is significant at the 10% level only. Both effects are tiny, so we conclude that the store does not condition prices on purchase histories.

Table A.3: Price dispersion within UPC-days

	Loyalty card and quantity		Purchase histories	
	(1)	(2)	(3)	(4)
Loyalty card	−0.276*** (0.102)			
Quantity		−0.105*** (0.017)		
# Shopping months			−0.00111 (0.00089)	
Log annual expenditure				0.004* (0.002)
UPC-day fixed effects	Yes	Yes	Yes	Yes
Sample	Mixed	All	Single unit	Single unit
Observations	382,300	1,605,379	1,026,682	1,026,682
R ²	0.38	0.40	0.73	0.73

Notes: The dependent variable is the unit price in dollars. All samples restrict to UPC-days with at least two transactions. “All” includes every transaction, by cardholders and anonymous shoppers. “Mixed” keeps the UPC-days where both cardholders and anonymous shoppers purchase. “Single unit” keeps the single-unit purchases of cardholders. “# Shopping months” is the number of months in which the consumer made at least one purchase at the store in 2018, and “Log annual expenditure” is the logarithm of the consumer’s total expenditure at the store over 2018. R² includes the UPC-day fixed effects. Standard errors in parentheses. Significance: * < 0.10; ** < 0.05; *** < 0.01.

B Estimates of Preference Parameters

B.1 Shopping occasion probabilities

Table B.1 presents the parameter estimates for the marginal distribution of the shopping occasion probabilities.

Table B.1: Estimated distribution of shopping occasion probability (ρ)

Latent class	ψ_u		ρ_u	
	Estimate	Std. err.	Estimate	Std. err.
1. Occasional shoppers	0.587	(0.005)	0.094	(0.002)
2. Regular shoppers	0.247	(0.003)	0.503	(0.004)
3. Frequent shoppers	0.167	(0.003)	0.942	(0.002)
Observations	20,748			

Notes: Parameters estimated from a three-class finite mixture model using our sample of 20,748 cardholders. Within class u , the number of months with a shopping occasion is distributed as a Binomial(12, ρ_u), truncated at zero. The shopping occasion probabilities ρ_u are estimated using a logit link function. The ψ_u are the class probabilities. Standard errors obtained using the delta method.

B.2 Supports and correlations of preference parameters

Table B.2 reports, for each category, the estimated supports of the preference parameters and the correlations between the parameters across consumer types. Consumers with stronger tastes for a category tend to be also more price sensitive: the price coefficient and the category taste are negatively correlated in 20 of the 24 categories. Frequent shoppers tend to have stronger category tastes and to be less price sensitive. The shopping occasion probability is positively correlated with the category taste in 22 categories and with the price coefficient in 18. The correlation between the own price elasticity and the category taste in dollars, in contrast, has no systematic sign. It is negative in 13 categories and positive in 11.

B.3 Distributions of price elasticities and category tastes

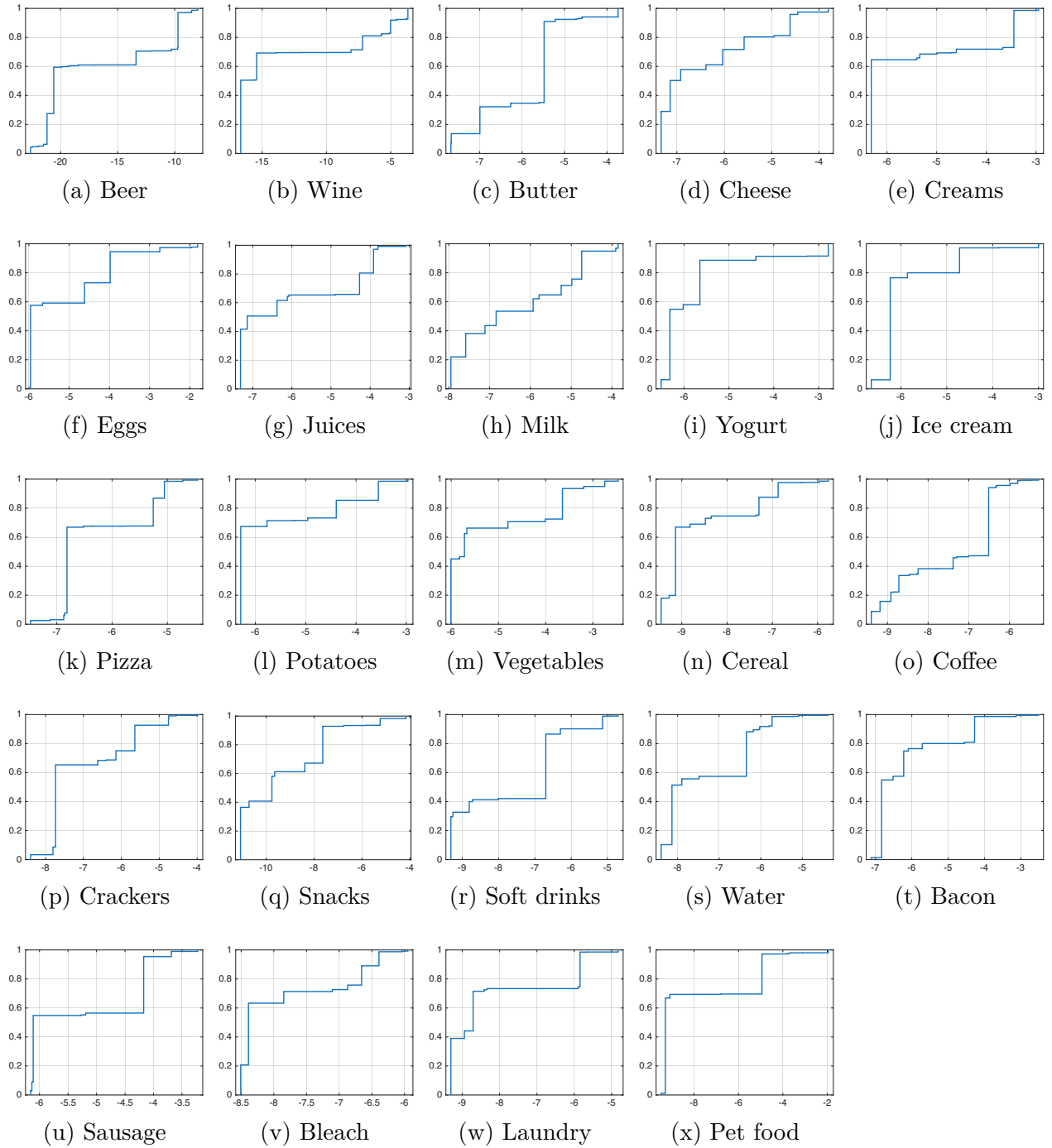
Figure B.1 and Figure B.2 show the distributions across consumer types of own price elasticities and of category tastes, expressed in dollars, in each category.

Table B.2: Parameter supports and correlations across consumer types

Category	α_d support		β_d support		Correlations			
	Lower	Upper	Lower	Upper	(α, β)	(α, ρ)	(β, ρ)	$(e, \beta/ \alpha)$
Alcohol								
Beer	-1.353	-0.348	-1.591	2.455	-0.597	0.204	-0.537	-0.652
Wine	-0.968	-0.076	-12.077	4.524	0.272	0.375	-0.047	0.241
Dairy								
Butter/margarine	-1.867	-0.925	-0.629	3.783	-0.478	-0.112	0.423	-0.339
Cheese	-0.936	-0.560	-4.908	7.342	-0.238	0.126	0.372	0.159
Creams	-1.327	-0.615	-1.152	8.170	0.199	0.111	0.094	0.427
Fresh eggs	-1.681	-0.689	-0.707	3.055	-0.295	0.093	0.393	0.272
Refr. juices	-1.249	-0.679	-7.410	11.979	-0.089	0.240	0.232	0.058
Milk	-1.368	-0.776	-1.331	3.626	-0.494	0.092	0.339	-0.042
Yogurt	-1.316	-0.538	-0.626	4.539	-0.336	-0.121	0.366	-0.243
Frozen food								
Ice cream	-1.047	-0.390	-0.353	3.515	-0.284	0.124	0.377	-0.113
Fz. pizza	-0.946	-0.594	-0.791	21.904	0.148	-0.070	0.306	0.375
Fz. potatoes	-1.395	-0.704	-0.418	4.434	-0.015	0.282	0.323	0.162
Fz. vegetables	-1.009	-0.475	-0.750	2.751	-0.335	0.315	0.197	-0.242
Grocery								
Cereal	-1.311	-0.952	-10.290	4.159	-0.137	0.280	0.166	0.157
Coffee	-0.870	-0.501	-1.317	5.817	-0.497	-0.214	0.248	-0.428
Crackers	-1.657	-1.036	-0.384	2.881	-0.074	0.109	0.496	0.190
Snacks	-1.380	-0.951	-7.604	2.868	-0.294	0.249	0.285	0.098
Soft drinks	-0.869	-0.535	-0.776	4.621	-0.484	0.022	0.326	-0.125
Water	-0.992	-0.576	-0.753	4.757	-0.302	-0.009	0.341	-0.080
Packaged meat								
Bacon	-0.714	-0.269	-0.687	2.940	-0.213	0.266	0.191	-0.176
Dinner sausage	-0.726	-0.390	-5.992	9.777	-0.449	0.002	0.378	-0.296
Taxable grocery								
Bleach	-1.696	-1.280	-2.156	9.068	0.082	0.386	0.475	0.300
Laundry detergent	-1.035	-0.617	-1.099	18.222	-0.471	0.112	0.007	-0.480
Pet food	-0.754	-0.032	-1.135	7.461	-0.258	-0.035	0.159	-0.375

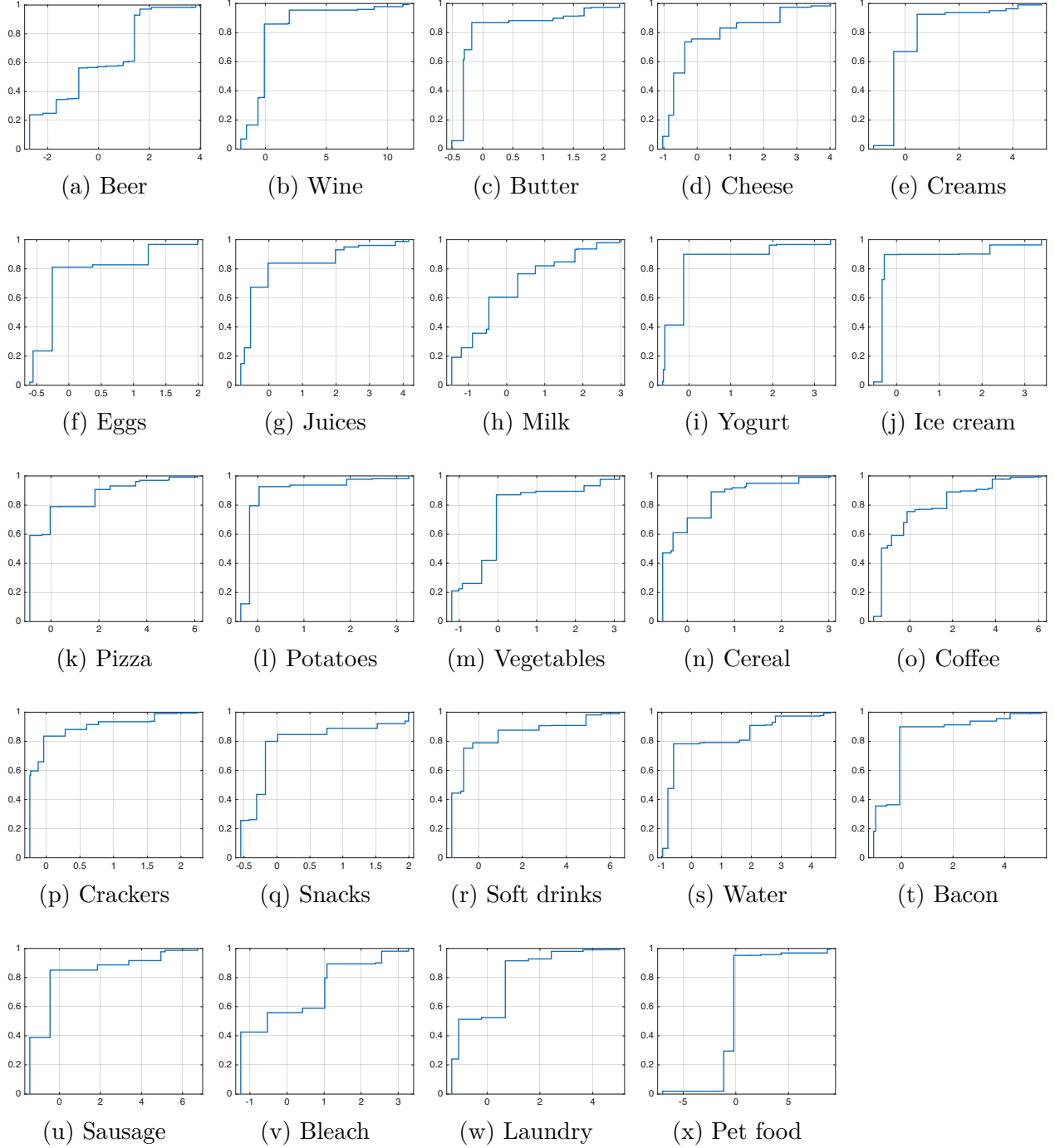
Notes: The support of α_d is reported at its estimated location; the support of β_d is reported under the mean zero normalization. Correlations are weighted by the estimated type probabilities ϕ_d ; e denotes the type level own-price elasticity and $\beta/|\alpha|$ the category taste in dollars.

Figure B.1: Distribution of own-price elasticities across consumers



Notes: Each panel is the CDF of the type level own-price elasticity in one category, weighted by the estimated type probabilities ϕ_d : the sales weighted average over products and months of $\alpha_d p_{jt}(1 - \mathbb{P}_{jt}(\theta_d, \delta_t))$, evaluated at observed prices; the weights are common to all types within a category. Distributions are winsorized at the 0.5th and 99.5th weighted percentiles; axis ranges differ across panels.

Figure B.2: Distribution of category tastes across consumers, in dollars



Notes: Each panel is the CDF of the type level category taste in one category, weighted by the estimated type probabilities ϕ_d and expressed in dollars: the deviation of β_d from the category mean, divided by $|\alpha_d|$. A value of one means the consumer values a purchase occasion in the category one dollar more than a consumer with the mean category taste and the same price sensitivity. Distributions are winsorized at the 0.5th and 99.5th weighted percentiles; axis ranges differ across panels.

C Additional Results on the Counterfactuals

C.1 Computational method

The dimensionality problem. The simulation of personalized prices relies on computing the posterior type probabilities, $\phi_{d|\mathbf{h}_\ell^\tau}$, and solving the first-order conditions (7) for each purchase history $\mathbf{h}_\ell^\tau$ of length τ . In each period, a consumer either has no shopping occasion, or chooses among the J products and the outside option, so the number of possible purchase histories grows as $(J + 2)^\tau$. This leads to a dimensionality problem: with $J = 5$ products, a twelve-month history already generates more than 10^{10} possible histories. Enumerating all purchase histories is therefore computationally prohibitive for the 12- to 36-month histories we study.

The problem becomes tractable once we observe that a purchase history affects the seller’s problem only through the posterior distribution it induces: two histories that lead to the same posterior lead to the same personalized prices, profits, and consumer surplus. We show below that the posterior admits a low-dimensional sufficient statistic, $(K, \mathbf{\Lambda})$, where K counts shopping occasions and $\mathbf{\Lambda}$ accumulates choice log-likelihood ratios. Instead of enumerating histories, we track the model-implied probability distribution of this statistic on a sparse grid and solve the pricing problem once per grid point.

A sufficient statistic for purchase histories. The estimated type grid combines a set of distinct preference parameters $\theta = (\alpha, \beta)$ with the three support points of the shopping occasion probability ρ , so each preference pair appears in three types. Let $\theta^{(1)}, \dots, \theta^{(A)}$ denote the distinct preference parameter pairs and let $a(d)$ index the pair of type d , that is, $\theta_d = \theta^{(a(d))}$. With discrete types, the likelihood in equation (5) factors into a shopping occasion component and a choice component,

$$f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell^\tau|\Theta_d) = \underbrace{\rho_d^K (1 - \rho_d)^{\tau-K}}_{\text{Shopping occasions}} \times \underbrace{\prod_{t=1}^{\tau} \prod_{j=0}^J \mathbb{P}_{jt}(\theta^{(a(d))})^{Y_{\ell jt}}}_{\text{Choices}}, \quad (15)$$

where $K = \sum_{t=1}^{\tau} \sum_{j=0}^J Y_{\ell jt}$ is the number of shopping occasions in the history. The first component depends on the history only through K , and the second component depends on the type only through its preference pair $\theta^{(a(d))}$.

Because Bayes’s rule involves only likelihood ratios, we can normalize the choice component by the choice likelihood of a reference pair $\theta^{(1)}$, which we take to be the most price-sensitive

one.⁴⁸ The choice information contained in a purchase history is then fully summarized by $A - 1$ cumulative log-likelihood ratios,

$$\Lambda_a(\mathbf{h}_\ell^\tau) = \sum_{t=1}^{\tau} \sum_{j=0}^J Y_{\ell jt} \cdot \ln \left(\frac{\mathbb{P}_{jt}(\theta^{(a)})}{\mathbb{P}_{jt}(\theta^{(1)})} \right), \quad a = 2, \dots, A,$$

with the normalization $\Lambda_1 = 0$. Substituting the factorization (15) into Bayes's rule, the reference choice likelihood cancels between the numerator and the denominator, and the posterior takes the closed form

$$\phi_{d|\mathbf{h}_\ell^\tau} = \frac{\phi_d \cdot \rho_d^K (1 - \rho_d)^{\tau-K} \cdot \exp(\Lambda_{a(d)})}{\sum_{d'=1}^D \phi_{d'} \cdot \rho_{d'}^K (1 - \rho_{d'})^{\tau-K} \cdot \exp(\Lambda_{a(d')})}. \quad (16)$$

The posterior, and therefore every object computed from it, depends on the purchase history only through the statistic $(K, \mathbf{\Lambda})$, where $K \in \{0, 1, \dots, \tau\}$ and $\mathbf{\Lambda} = (\Lambda_2, \dots, \Lambda_A)$. Across our product categories, the number of distinct preference pairs ranges from 8 to 25, so $\mathbf{\Lambda}$ has at most 24 components and the statistic $(K, \mathbf{\Lambda})$ has at most 25, regardless of the number of products and the length of the history.

Two remarks are in order. First, the increments in Λ_a depend on the prices in each period because purchasing under different price levels carries different information. For example, buying when the price is high is better evidence of low price sensitivity than when the price is low. This is why a simple purchase count does not summarize the choice component and Λ_a must be accumulated period by period. Second, as discussed in Section 2, the correlation between ρ and (α, β) in the prior implies that K is informative about preferences and $\mathbf{\Lambda}$ about the shopping occasion probability. The posterior (16) does not factor into separate beliefs about ρ and θ , so we always carry the joint posterior over the D types.

Computing the distribution of beliefs. Aggregating outcomes across consumers requires the model-implied probability of reaching each posterior belief. We call each value of the statistic $(K, \mathbf{\Lambda})$ a belief node, and denote by $\mu_\tau(K, \mathbf{\Lambda})$ its probability mass, the sum of the unconditional probabilities $f_{\mathbf{h}_\ell}(\mathbf{h}_\ell^\tau)$ of the histories of length τ that map to it. We compute μ_τ with an exact forward recursion rather than by simulating histories. The recursion starts from a unit mass at $(0, \mathbf{0})$, the empty history. In each period t , each belief node passes its mass to the nodes that the next outcome leads to, in proportion to the probability of each

⁴⁸The choice of the reference pair is immaterial: the posterior is invariant to it.

outcome under the node’s posterior:

$$\begin{aligned} \text{No shopping occasion:} \quad & \sum_{d=1}^D \phi_{d|(K, \mathbf{\Lambda})} \cdot (1 - \rho_d), \quad (K, \mathbf{\Lambda}) \text{ unchanged,} \\ \text{Shopping occasion, choice } j: \quad & \sum_{d=1}^D \phi_{d|(K, \mathbf{\Lambda})} \cdot \rho_d \cdot \mathbb{P}_{jt}(\theta_d), \quad (K, \mathbf{\Lambda}) \mapsto (K + 1, \mathbf{\Lambda} + \mathbf{\Lambda}_t(j)), \end{aligned}$$

where $\mathbf{\Lambda}_t(j)$ collects the period- t increments $\ln \left[\mathbb{P}_{jt}(\theta^{(a)}) / \mathbb{P}_{jt}(\theta^{(1)}) \right]$, for $a = 2, \dots, A$. Histories that reach the same belief node pool their probability masses automatically, which is the source of the computational gains.

The count K is discrete and carried exactly. The $\mathbf{\Lambda}$ coordinates are carried on an adaptive sparse support. After each period, nodes that share the same K and whose $\mathbf{\Lambda}$ coordinates agree within a tolerance of 10^{-4} are merged into their probability-weighted centroid. The merge conserves total probability exactly and approximates $\mathbf{\Lambda}$ up to the tolerance level. Whenever the number of nodes exceeds a cap (12,000 nodes for $\tau \leq 12$, increasing linearly to 360,000 at $\tau = 36$), the merge tolerance is doubled until the count falls below the cap.⁴⁹ The cap binds in every category beyond the first few months, so the realized tolerance is far coarser than its initial value, and the merge combines histories whose choice evidence is similar rather than identical. The accuracy checks reported below measure the consequences of this approximation.

Histories beyond the estimation window. The recursion uses the observed monthly prices and product qualities $\{p_{jt}, \delta_{jt}\}$ over the 12 months of the sample, in their observed order. For horizons $\tau > 12$, each additional month is assigned the prices and the qualities of one of the 12 observed months, drawn at random with replacement. We draw a single sequence of 24 months, which we attach to our 12 months of observed histories to form a 36-month sequence, and use its first τ entries at each horizon, such that outcomes vary across τ only through the accumulation of information and not through resampled environments.

Calculating profits, consumer surplus, and the information index. Let r index the terminal belief nodes and write $\phi_{d|r}$ and $\mu_\tau(r)$ for the associated posterior and probability mass. At each node, we solve the first-order conditions (7), with the posterior defined in (16) in place of the history-specific posterior, using the fixed-point algorithm proposed by

⁴⁹Water uses the 360,000 cap at every horizon. With the smaller cap, the merge was too coarse in this category to pass the accuracy checks reported below.

Morrow and Skerlos (2011). Denoting the resulting prices by $p_j(r)$, expected profits under personalized pricing are

$$\pi^H = \sum_r \sum_{j=1}^J (p_j(r) - c_j) \cdot \left[\sum_{d=1}^D \mathbb{P}_j(\theta_d) \rho_d \phi_{d|r} \right] \cdot \mu_\tau(r) \cdot M,$$

where $\mu_\tau(r) \cdot M$ is the expected number of consumers whose purchase history reaches node r .

Consumer surplus is aggregated in the same way using the log-sum formula in equation (2):

$$CS = - \sum_r \sum_{d=1}^D \left(\frac{\rho_d}{\alpha_d} \ln \left(1 + \sum_{j=1}^J \exp(\delta_j + \beta_d + \alpha_d p_j(r)) \right) \phi_{d|r} \right) \cdot \mu_\tau(r) \cdot M.$$

The aggregate information index is computed as:

$$\mathcal{I}_\eta(\tau) = \sum_r \mathcal{I}_\eta(r) \cdot \mu_\tau(r).$$

Distributional results, such as the dispersion of personalized prices or the change in consumer surplus, treat the node-level outcomes as a discrete distribution with weights $\mu_\tau(r)$.

Profits and consumer surplus under uniform pricing and perfect price discrimination, whose prices solve equations (3) and (8) respectively (and do not depend on beliefs), are evaluated on the same belief nodes with the same weights, which ensures that the three pricing regimes are compared under exactly the same distribution of types.

Finally, within the sample ($\tau \leq 12$), each observed consumer can be assigned to the belief node reached by their actual purchase history, which we use for the consumer-level results that combine information across categories.

Accuracy checks. The recursion implies two aggregation identities that we use as diagnostics. First, averaging the posteriors over belief nodes must recover the prior, $\sum_r \phi_{d|r} \cdot \mu_\tau(r) = \phi_d$. Second, by the law of total variance, the aggregate information index can be computed equivalently from the expected posterior variance, as in Section 5.2, or as the variance of the node-level posterior means divided by the prior variance. Both identities hold exactly in the absence of node merging. Therefore, any deviation from these identities measures the approximation error introduced by node merging. Across all categories, the prior is recovered with a maximum error of 1.4×10^{-3} at $\tau = 12$ and 2.0×10^{-3} at $\tau = 36$, and the two computations of the information index agree within 6×10^{-3} and 1.2×10^{-2} respectively.

Learning while price discriminating. In our benchmark analysis, purchase histories accumulate under uniform pricing, that is, the seller charges the uniform prices during the observation window and personalizes prices only in period $\tau + 1$, as described in Section 2. Because marginal costs are recovered by inverting the uniform-pricing first-order conditions (3) at the observed prices, the observed monthly prices are the model’s uniform optimal prices, and the recursion evaluates the choice probabilities and the increments $\Lambda_t(j)$ at those prices.

We also consider a variant in which the seller learns about consumer types while it sets personalized prices. Only product qualities δ_{jt} are taken from the observed months, and prices are determined by the seller’s beliefs at each node. In the first period, the belief is equal to the prior, so every consumer faces the same prices, which coincide with our uniform pricing benchmark. In every subsequent period, the personalized prices at each belief node solve the first-order conditions (7) at the posterior the seller holds at the beginning of the period. As throughout the paper, the seller prices myopically and consumers do not anticipate future prices.

The statistic (K, Λ) remains sufficient under repricing. Prices are a deterministic function of the belief nodes, and histories that reach the same node face the same prices, so they accumulate in identical increments. The forward recursion carries over unchanged, though each node now requires an optimal price solve in every period rather than a single solve at the terminal nodes. Because shopping occasions do not respond to prices, repricing leaves the distribution of K exactly unchanged, which we use as an additional check on the implementation.

Incomplete histories. The coarser information structures of Table 3 use the same recursion, and only its inputs change. When the seller does not observe which product is chosen, the J products are collapsed into a single purchase alternative. When it does not observe store visits, the seller observes each month either a purchase of product j or no purchase (reflecting either a shopping occasion without a purchase or no shopping occasion). For type d , these outcomes have probabilities $\rho_d \mathbb{P}_{jt}(\theta^{(a(d))})$ and $1 - \rho_d + \rho_d \mathbb{P}_{0t}(\theta^{(a(d))})$, which combine the shopping occasion probability with the choice probabilities. The recursion then treats every month as having one of these $J + 1$ outcomes. Because these probabilities differ across types that share a preference pair, Λ has one coordinate per type rather than per preference pair. When the seller observes only store visits, the statistic reduces to K , and the belief nodes

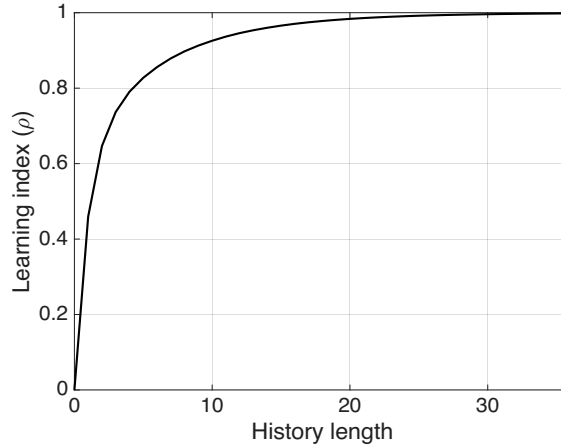
and their weights are available in closed form.

C.2 Additional results

C.2.1 Information acquisition

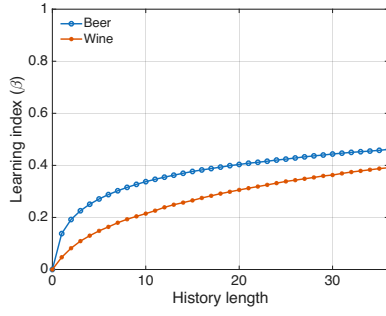
Figure C.1 shows the learning index for the shopping occasion probability ρ , and Figure C.2 the index for the category taste β in every category, grouped by department. We show a single curve for ρ because the distribution of ρ is identical across categories, and store visits carry almost all the information about ρ . Choices are indirectly informative about ρ through its correlation with preferences, which makes the index differ slightly across categories, but the resulting differences are at most 0.02 in the first month and below 0.01 from the third month on.

Figure C.1: Learning index for the shopping occasion probability (ρ)

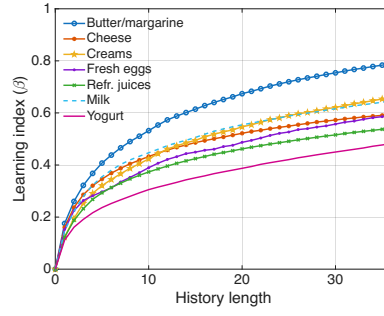


Notes: History length in months. The figure shows a single category; the index is nearly identical across categories.

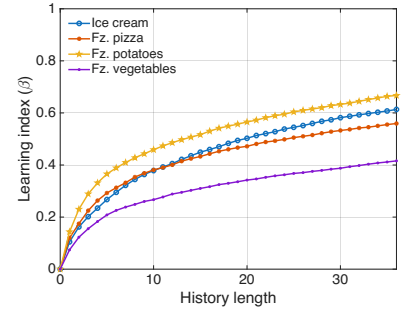
Figure C.2: Learning index for the product category taste parameter (β)



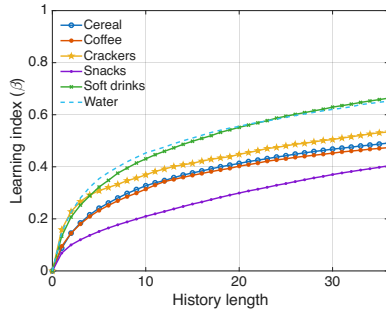
(a) Alcohol



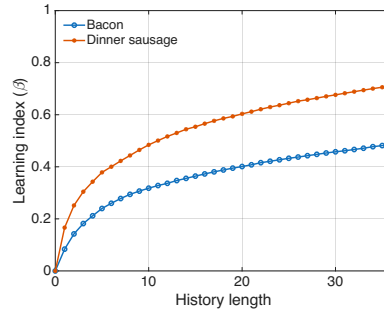
(b) Dairy



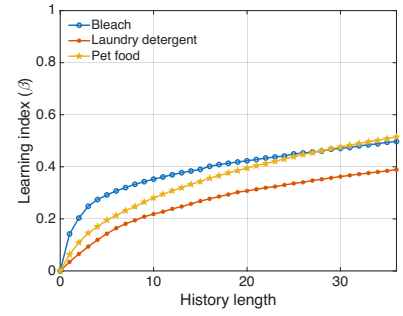
(c) Frozen food



(d) Grocery



(e) Packaged meat



(f) Taxable grocery

Notes: History length in months. Each figure shows one product department, with separate lines for the categories within it.

C.2.2 Dispersion of personalized prices

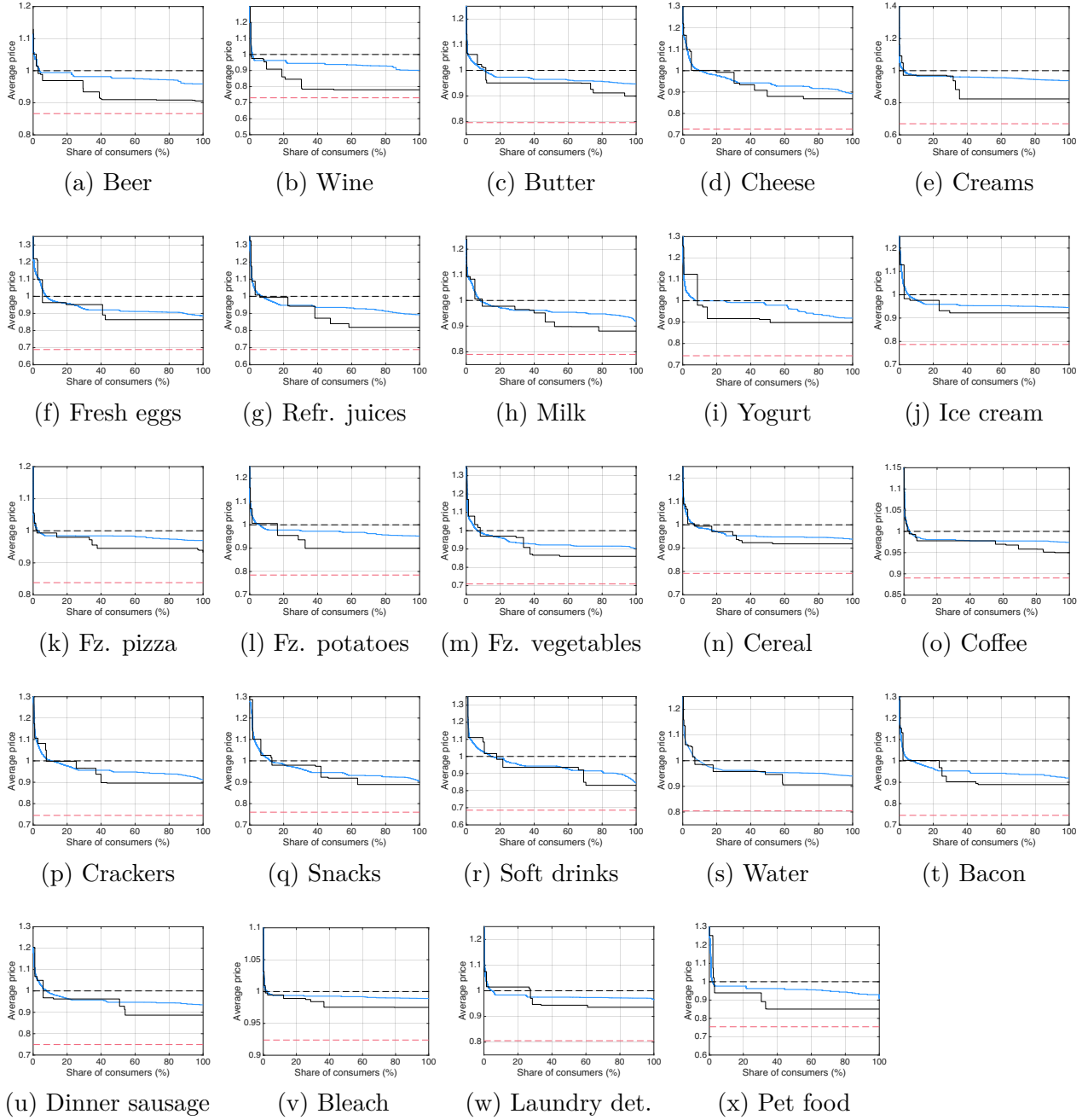
Table C.1 reports the convergence of surcharges and discounts toward perfect discrimination at 3, 12, and 36 months. Figure C.3 shows the distribution of prices with 12-month histories for all 24 categories, constructed as in Figure 5.

Table C.1: Convergence of surcharges and discounts toward perfect discrimination

Category	Share of the initial price gap closed (%)					
	3 months		12 months		36 months	
	Surcharges	Discounts	Surcharges	Discounts	Surcharges	Discounts
Alcohol						
Beer	63.7	17.2	84.2	30.7	94.5	40.9
Wine	73.6	23.1	80.4	33.9	85.6	40.2
Dairy						
Butter/margarine	45.6	41.2	63.1	58.7	78.7	72.6
Cheese	71.2	53.2	68.2	69.3	77.9	80.6
Creams	39.4	22.8	46.9	33.4	50.5	45.6
Fresh eggs	58.2	54.4	75.5	74.7	87.0	84.4
Refr. juices	76.0	40.7	84.1	54.7	95.2	66.2
Milk	81.9	44.5	88.4	57.6	90.9	68.3
Yogurt	20.4	21.7	29.6	35.2	50.4	55.5
Frozen food						
Ice cream	68.0	43.5	76.8	65.7	87.2	77.6
Fz. pizza	98.4	32.4	92.0	46.3	93.0	52.2
Fz. potatoes	-9.2	26.7	-7.5	37.0	16.4	47.4
Fz. vegetables	59.7	47.3	59.8	66.5	63.3	77.6
Grocery						
Cereal	88.2	52.3	93.7	72.2	95.0	82.7
Coffee	98.8	47.1	129.1	69.4	111.2	79.2
Crackers	56.1	48.0	70.4	64.8	79.4	76.4
Snacks	66.8	61.8	78.0	77.8	87.7	86.8
Soft drinks	51.4	52.7	67.1	70.9	82.0	83.5
Water	64.2	49.4	81.0	65.4	95.3	75.3
Packaged meat						
Bacon	-29.4	37.3	-7.1	57.2	26.5	73.1
Dinner sausage	63.5	44.8	77.5	61.9	91.8	73.1
Taxable grocery						
Bleach	129.8	32.7	92.0	41.4	90.7	46.8
Laundry detergent	-46.8	25.5	-35.2	44.5	-53.3	53.6
Pet food	20.3	22.2	44.4	35.9	64.6	48.1
Median across categories	63.6	42.4	76.2	58.2	86.3	72.8

Notes: For each category and history length, we compute the area between the uniform price line and the personalized price curve, relative to the area between the uniform price line and the perfect discrimination curve, separately over the segment where perfect discrimination prices exceed the uniform price (surcharges) and the segment where they lie below it (discounts), in percent. Negative values indicate that, on that segment, personalized prices lie on average on the other side of the uniform price than the perfect discrimination prices. Values above 100 indicate that personalized prices overshoot the perfect discrimination prices on that segment. Histories are weighted by their model-implied probability mass. Consumers are ranked separately under each regime, so the shares compare distributions of prices, not the prices of the same consumers.

Figure C.3: Distribution of prices, 12-month histories, all categories

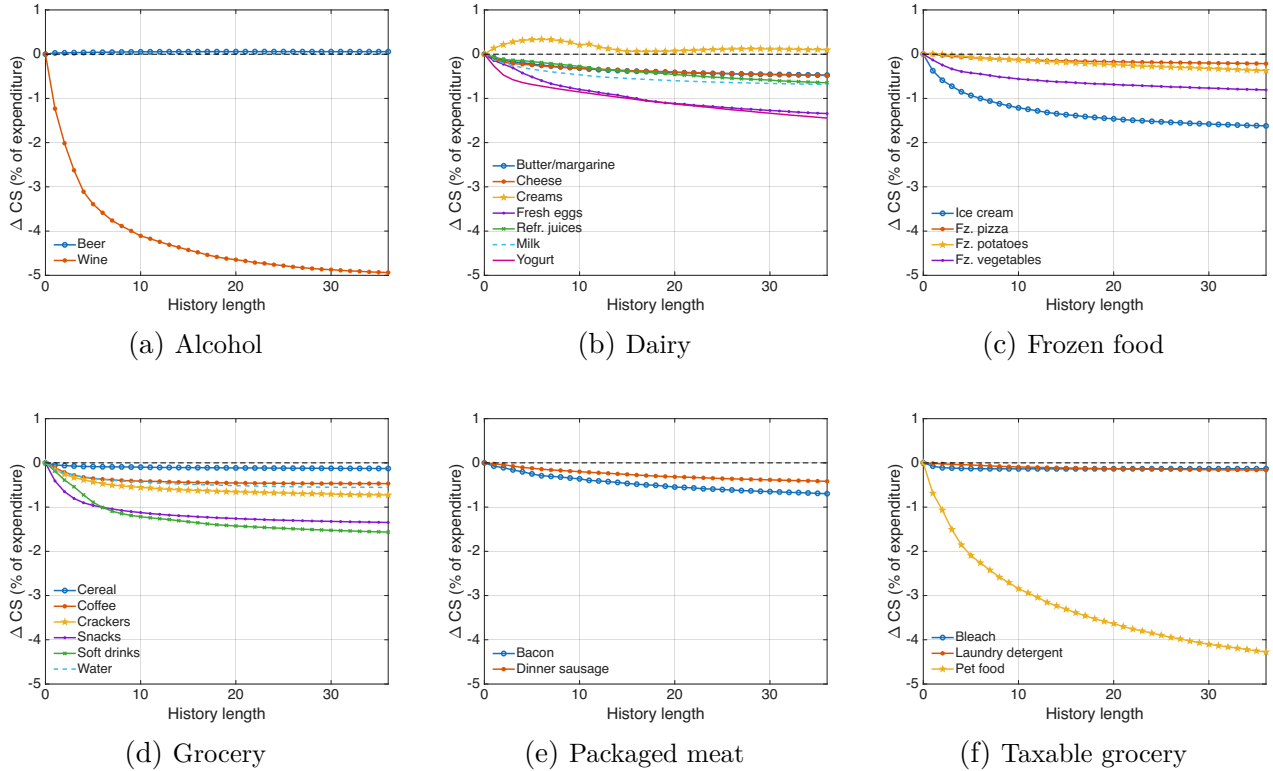


Notes: Average personalized (blue) and perfect price discrimination (black) prices normalized by the average uniform price, across the consumer distribution, for 12-month histories; consumers ranked from highest-paying to lowest-paying. The dashed black line is the uniform price and the dashed red line the average marginal cost, normalized by the average uniform price. In 19 categories, the vertical axis is cut below the top of the perfect discrimination curve; the share of consumers priced above the axis is at most about 1% under either pricing regime.

C.2.3 Consumer surplus

Figure C.4 depicts the change in consumer surplus by category, in percentage of expenditure, as a function of the histories length. Consumer surplus falls in 21 of the 24 categories at every history length, and in frozen potatoes from the second month on. Only beer and creams gain at every history length, by at most 0.06% and 0.34% of expenditure, respectively. The losses are small relative to spending: with 12-month histories, the median category loses 0.4% of the average expenditure, 19 categories lose less than 1%, and the largest losses are 4.2% in wine and 3.0% in pet food, followed by ice cream, soft drinks, and snacks between 1.2 and 1.3%. In cereal, 93% of consumers gain, the median consumer gains 0.8% of average expenditure and the consumer at the 90th percentile gains 6.4%, yet the mean change is negative because the consumer at the 1st percentile loses 63% of the average expenditure. In eight categories the hardest-hit percent of consumers lose more than the average consumer spends in the category.

Figure C.4: Change in consumer surplus under personalized pricing, by category



Notes: Each line represents the evolution of the change in aggregate consumer surplus, from uniform to personalized pricing, in percentage of the average category expenditure under uniform pricing.

C.2.4 Profitability of personalized pricing

Table C.2 reports, by category, the profit gain under perfect price discrimination and the share of this gain that personalized pricing captures with histories of 3, 12, and 36 months.

Table C.2: Profit gains relative to perfect price discrimination, by category

Category	Δ Profit (%)	Share of the perfect gain captured (%)		
	Perfect	3-month	12-month	36-month
Alcohol				
Beer	6.7	35.9	66.6	84.8
Wine	23.7	38.0	73.0	90.6
Dairy				
Butter/margarine	6.0	33.0	65.4	86.7
Cheese	10.0	38.1	65.6	82.8
Creams	5.5	19.3	40.6	65.4
Fresh eggs	10.5	32.6	62.5	85.2
Refr. juices	10.9	35.6	63.2	82.3
Milk	10.1	38.4	67.8	86.6
Yogurt	12.1	28.0	52.2	72.3
Frozen food				
Ice cream	10.6	33.3	65.9	87.2
Fz. pizza	4.8	27.6	60.8	81.2
Fz. potatoes	6.4	24.3	50.9	72.5
Fz. vegetables	10.1	27.7	53.6	72.1
Grocery				
Cereal	7.0	39.0	68.0	84.9
Coffee	7.4	38.3	71.7	88.7
Crackers	9.3	32.3	63.6	82.5
Snacks	14.4	43.1	72.3	88.6
Soft drinks	14.0	33.7	68.3	88.8
Water	9.7	40.8	70.0	87.6
Packaged meat				
Bacon	8.9	27.3	54.4	76.2
Dinner sausage	6.2	32.1	61.7	84.0
Taxable grocery				
Bleach	5.0	45.2	77.9	91.8
Laundry detergent	4.4	26.4	55.7	78.8
Pet food	22.4	24.5	56.6	80.0
All categories (aggregate)	10.8	35.3	65.9	85.2

Notes: The first column reports the percentage change in expected profits from uniform pricing to perfect price discrimination. The last three columns express the profit gain from personalized pricing with purchase histories of 3, 12, and 36 months as a percentage of this perfect discrimination gain. The bottom row aggregates profits across the 24 categories.

Table C.3 reports the share of the profit gain from personalized pricing earned on consumers grouped by the number of purchases in their 12-month history.

Table C.3: Profit gains from personalized pricing by number of purchases in the history

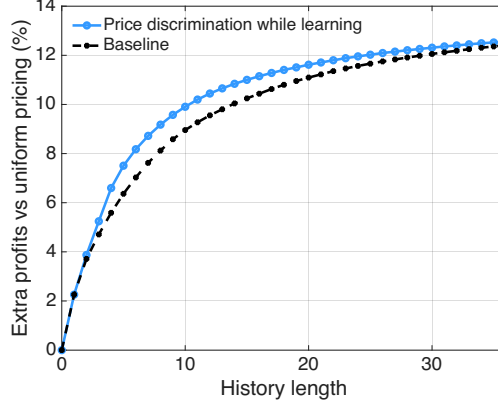
Category	# purchases = 0		1 ≤ # purchases ≤ 3		# purchases ≥ 4	
	Consumers (%)	Gain (%)	Consumers (%)	Gain (%)	Consumers (%)	Gain (%)
Alcohol						
Beer	89.2	28.8	9.4	14.7	1.4	56.5
Wine	92.0	10.8	7.0	8.6	0.9	80.6
Dairy						
Butter/margarine	63.7	14.7	28.7	17.3	7.6	67.9
Cheese	50.7	20.9	34.1	15.6	15.2	63.5
Creams	85.1	18.7	12.7	6.1	2.1	75.2
Fresh eggs	59.8	21.2	29.7	10.5	10.5	68.3
Refr. juices	68.8	19.9	23.9	12.8	7.3	67.3
Milk	52.0	18.3	32.8	15.4	15.2	66.3
Yogurt	83.2	20.3	14.7	10.9	2.1	68.7
Frozen food						
Ice cream	75.0	11.9	21.4	15.5	3.6	72.5
Fz. pizza	87.2	16.5	11.2	9.5	1.6	74.0
Fz. potatoes	78.9	19.2	17.9	11.0	3.2	69.8
Fz. vegetables	66.7	19.5	26.2	10.1	7.1	70.4
Grocery						
Cereal	69.3	24.3	24.5	9.0	6.2	66.6
Coffee	83.2	15.7	14.5	10.1	2.3	74.2
Crackers	61.8	21.8	29.8	10.3	8.4	67.8
Snacks	47.9	16.0	36.3	14.0	15.8	70.0
Soft drinks	42.5	14.9	38.0	14.8	19.5	70.3
Water	62.9	18.9	28.3	13.6	8.8	67.6
Packaged meat						
Bacon	71.1	22.8	23.6	8.3	5.3	68.9
Dinner sausage	68.3	19.2	24.5	12.9	7.2	68.0
Taxable grocery						
Bleach	87.5	13.1	10.9	8.2	1.6	78.7
Laundry detergent	88.8	17.5	10.1	24.0	1.1	58.5
Pet food	93.4	6.5	5.8	20.5	0.8	73.0
All categories (pooled)	72.1	17.3	21.5	13.5	6.5	69.2

Notes: 12-month histories of the observed consumers, grouped by the number of purchases in the category. “Consumers” is the group’s share of the category’s consumers and “Gain” is the group’s share of the category’s profit gain from personalized pricing, both in percent. The last row pools the 24 categories over the 20,748 cardholders.

C.2.5 Price discrimination while learning

Figure C.5 shows the evolution of the profit gain under our baseline personalized prices against the personalized prices that arise if the seller price discriminates during the learning process. The profit gains along this path are at least as large as in the baseline at every history length. They reach 5.2% against 4.7% with 3-month histories, 10.4% against 9.6% with 12-month histories, and 12.6% against 12.4% with 36-month histories.

Figure C.5: Profit gains when the seller price discriminates while learning in soft drinks



Notes: Percent change in profits from uniform to personalized pricing, by history length, when the seller holds the uniform price during the learning phase (baseline) and when it reprices every month on its current beliefs.

C.2.6 Learning across categories

The seller updates its beliefs about a consumer's type in a category from that consumer's history in that category alone, although types may be correlated across categories. Relaxing this assumption for all 24 categories is not feasible. The seller's prior would have to assign a probability to every combination of types across the 24 categories. With 24 to 75 types per category, the number of combinations is far too large to estimate a weight for each of them. Restricting the dependence with a parametric form would give up the unrestricted type distribution on which our estimation relies. In addition, the set of possible joint histories multiplies with the number of products in each category, so the recursion of Appendix C.1, which tracks the seller's belief states month by month, can handle only short histories even for two categories.

We measure the value of pooling histories for two pairs of frequently co-purchased categories, milk and cereal, and snacks and soft drinks. For each pair, we estimate the joint distribution of the two categories' types holding fixed the estimated marginal distribution of types in each category. Denote the estimated types of category A by $\Theta_d^A = (\theta_d^A, \rho_d)$, $d = 1, \dots, D^A$, with weights ϕ_d^A , and likewise $\Theta_{d'}^B$ and $\phi_{d'}^B$ for $d' = 1, \dots, D^B$. Because the shopping occasion probability is common to all categories, a joint type is a pair (d, d') whose two components share the same value ρ_u . Conditional on the pair (d, d') , the probability of consumer i 's pooled history, with K_i store visits, is the product of the visit probabilities and of the choice

probabilities in the two categories, as in equation (9):

$$\mathcal{L}_i(d, d') = \rho_u^{K_i} (1 - \rho_u)^{T-K_i} \prod_{t=1}^T \left[\prod_{j=0}^{J^A} \mathbb{P}_{jt}^A(\theta_d^A)^{Y_{ijt}^A} \right] \left[\prod_{j=0}^{J^B} \mathbb{P}_{jt}^B(\theta_{d'}^B)^{Y_{ijt}^B} \right].$$

The joint weights $\pi_{dd'}$ solve

$$\begin{aligned} \max_{\pi \geq 0} \quad & \sum_{i=1}^M \ln \left(\sum_{(d, d')} \pi_{dd'} \mathcal{L}_i(d, d') \right) - M \ln \left(\sum_{u=1}^U \psi_u \left[1 - (1 - \rho_u)^T \right] \right) \\ \text{s.t.} \quad & \sum_{d': \rho_{d'} = \rho_d} \pi_{dd'} = \phi_d^A \quad \forall d, \quad \sum_{d: \rho_d = \rho_{d'}} \pi_{dd'} = \phi_{d'}^B \quad \forall d', \end{aligned}$$

where the second term of the objective conditions the likelihood on the consumer having visited the store at least once, as in the baseline estimation. The objective is concave in the weights and the constraints are linear, so any local maximum is the global one. With 36 milk types and 42 cereal types, the joint distribution has 429 free parameters, and 243 for snacks and soft drinks.

Types are correlated across categories: the correlation of price sensitivities is 0.24 between milk and cereal and 0.43 between snacks and soft drinks, and the correlation of category tastes is 0.54 and 0.48 respectively. This correlation goes well beyond what the common shopping occasion probability implies (at most 0.10).

Histories are generated by the estimated joint model at the observed prices. With 3-month histories, we enumerate every sequence of store visits and joint choices with its probability, so that the comparison involves no approximation. With 12-month histories, the number of joint belief states is beyond the recursion of Appendix C.1, so we simulate one million consumers from the joint distribution.⁵⁰

Table C.4 shows that pooling adds little to the gains: less than one percentage point with 3-month histories and at most 0.5 percentage point with 12-month histories, against gains of 4.7% to 10.4% from the category's own history. For these pairs, category-by-category learning is therefore a close approximation of what a seller pooling the two histories would achieve.

⁵⁰The simulated gains of the seller that prices on each category's own history reproduce those obtained by the recursion in Table C.2.

Table C.4: Profit gains from pooling purchase histories across two categories

Category	3-month histories			12-month histories		
	Own	Pooled	Gain	Own	Pooled	Gain
Milk (with cereal)	3.9	4.4	0.5	6.9	7.2	0.3
Cereal (with milk)	2.7	3.3	0.6	4.7	5.1	0.4
Snacks (with soft drinks)	6.2	7.0	0.8	10.4	10.8	0.5
Soft drinks (with snacks)	4.7	5.5	0.8	9.6	10.0	0.4

Notes: Percent change in profits from uniform to personalized pricing when the seller prices each category on the posterior from that category's own purchase history (Own) and on the posterior from the histories of both categories of the pair (Pooled). Gain is the difference in percentage points.

C.2.7 Learning from a misspecified prior

Consider a seller who updates from a prior $\tilde{f}_\Theta$ that differs from the population distribution f_Θ . The prior $\tilde{f}_\Theta$ puts weight γ on a uniform distribution over the type grid and weight $1 - \gamma$ on f_Θ . By Bayes's rule, equation (4) with $\tilde{f}_\Theta$ in place of f_Θ , the posterior is

$$\tilde{f}_{\Theta|\mathbf{h}_\ell}(\Theta_i|\mathbf{h}_\ell) = \frac{f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i) \cdot \tilde{f}_\Theta(\Theta_i)}{\sum_{i'} f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_{i'}) \cdot \tilde{f}_\Theta(\Theta_{i'})}.$$

The likelihood of the history under each type, $f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i)$, does not depend on the prior. By equation (5), it is determined by the realized choices and by the type's choice and visit probabilities at the prices faced by each consumer. The misspecified prior enters only through the weights $\tilde{f}_\Theta(\Theta_i)$. Suppose the consumer is of type Θ_{i^*} . Dividing the numerator and the denominator by the term of the true type, the posterior on the true type is

$$\tilde{f}_{\Theta|\mathbf{h}_\ell}(\Theta_{i^*}|\mathbf{h}_\ell) = \frac{1}{1 + \sum_{i \neq i^*} \frac{\tilde{f}_\Theta(\Theta_i)}{\tilde{f}_\Theta(\Theta_{i^*})} \cdot \frac{f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_i)}{f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell|\Theta_{i^*})}}. \quad (17)$$

The prior ratio does not depend on the history, and the likelihood ratio does not depend on the prior. As the history lengthens, the likelihood ratio of every type that behaves differently from the true type, in its choices or its shopping occasions, converges to zero. Each month, the observed outcome is on average more likely under the true type than under any other type, so the evidence against it accumulates month after month. Since the prior ratios are

fixed, the posterior on the true type converges to one regardless of the prior, provided the prior puts positive weight on the true type.

The computation requires no new recursion. Histories are generated at the observed prices, as in the baseline, so their probabilities do not depend on the seller’s prior. For each history, we recompute the seller’s posterior from the misspecified prior and solve for the personalized prices at this posterior. We then evaluate the profits earned at these prices with the posterior implied by the correct prior.

C.3 The seller’s incentive to improve learning

Setup. In the two-period model of Section 5.6, the environment is stationary. Product qualities and marginal costs are the same in both periods, so the uniform price $\mathbf{p}^U$ and the uniform profit π^U are the same in both. The two periods differ only in the seller’s information. Hence $\pi_1(\mathbf{p}^U) = \pi^U$. A seller without information would also earn π^U in the second period, so $\pi_2(\mathbf{p}_1) - \pi^U$ is the value of the information contained in the histories generated at $\mathbf{p}_1$. At uniform prices, it is 1.86% of π^U .

Computation. For a candidate price vector $\mathbf{p}_1$, we compute the distribution of one-period purchase histories and the posterior associated with each of them, using the recursion of Appendix C.1 for a single period. We then solve for the personalized prices of each history from equation (7), and $\pi_2(\mathbf{p}_1)$ is the expected personalized profit over histories. First-period prices are restricted to $c_j \leq p_{j1} \leq 2.2 p_j^U$ for every product j . The lower bound rules out selling below marginal cost. The upper bound is high enough to price a product out of the market.⁵¹ We maximize $\pi_2(\mathbf{p}_1)$ in every category by multistart Nelder–Mead within these bounds. Each search starts from the uniform price and 15 random points.⁵² As a check on the search, we compute $\pi_2(\mathbf{p}_1)$ on a grid in the two categories with the fewest products, fresh eggs and snacks. In both, the search attains at least the grid maximum. For the forward-looking seller, we maximize $\pi_1(\mathbf{p}_1) + \pi_2(\mathbf{p}_1)$ with the same bounds and the same multistart Nelder–Mead algorithm.

Table C.5 reports the results by category. For the best experiment, it gives the value of

⁵¹Re-optimizing with the upper bound raised to $5 p_j^U$ raises the maximal gain by less than 0.1% of uniform profit.

⁵²When these starts do not agree on the optimum, we run a second pass with 200 new starts. This occurs in eight categories.

experimentation and the change in first-period profit, both in percent of uniform profit, and the break-even horizon. For the forward-looking seller, it gives the largest deviation of its first-period prices from the uniform price and the change in its profit over the two periods. The break-even horizon of the best experiment is shortest in pet food, at 26 months, and exceeds 10 years in most categories. The forward-looking seller keeps every price within 0.5% of the uniform price in every category, and its two-period profit exceeds the myopic seller’s by at most 0.005%.

Longer experiments. In fresh eggs, the only category with two products, we let the experiment last two periods and maximize the expected third-period personalized profit over the pair of price vectors $(\mathbf{p}_1, \mathbf{p}_2)$. As in the two-period model, the environment is stationary, so the three periods share the same product qualities and marginal costs. Posteriors depend on the prices a consumer has faced but not on their order, so the objective is symmetric in $(\mathbf{p}_1, \mathbf{p}_2)$. We evaluate unordered pairs only, on a coarse grid, followed by a local refinement around its maximum. The best pair of price vectors charges the same prices in both periods, and the gain is 0.18% of the myopic seller’s third-period profit. Only 21.6% of consumers have two shopping occasions over the two periods, so they are the only ones who could reveal more by facing different prices in the two periods. The others visit at most once and generate their history at the prices of that period alone. Charging the best one-period experiment in both periods guarantees that each of them faces it, whichever period they visit.

Experimentation along the learning path. The seller may have more to gain from experimenting once it has started to learn, because it can then experiment differently on consumers with different purchase histories. We let the seller charge history-specific experimental prices for one period after three periods of learning under uniform pricing, in the three categories where the initial experiment gains the most (pet food, wine, and soft drinks) and in the two-product category, fresh eggs. For each history, we solve the one-period problem above with the seller’s posterior in place of the prior, with the same bounds and the same multistart Nelder–Mead algorithm. The benchmark is a myopic seller that charges each history its personalized price. Tailoring the experiment to existing histories makes it more attractive in three of the four categories, where it gains more and breaks even sooner than the initial experiment. Tailored experiments gain between 0.56% (fresh eggs) and 1.57% (wine) of uniform profit, while the seller forgoes 32% to 63% of one period’s profit. Recovering this

Table C.5: The value of price experimentation, by category

Category	Best experiment			Forward-looking seller	
	Second period (%)	First period (%)	Break-even (months)	Price deviation (%)	Two periods (%)
Alcohol					
Beer	0.09	-53.8	605	0.05	0.000
Wine	1.14	-65.4	57	0.33	0.003
Dairy					
Butter/margarine	0.40	-79.9	199	0.11	0.001
Cheese	0.32	-60.1	190	0.13	0.001
Creams	0.51	-90.1	178	0.03	0.000
Fresh eggs	0.13	-35.0	270	0.11	0.000
Refr. juices	0.23	-33.9	148	0.23	0.000
Milk	0.36	-58.0	160	0.16	0.001
Yogurt	0.59	-34.2	58	0.18	0.001
Frozen food					
Ice cream	0.31	-42.3	135	0.14	0.000
Fz. pizza	0.06	-46.0	736	0.10	0.000
Fz. potatoes	0.07	-49.6	698	0.23	0.000
Fz. vegetables	0.25	-30.3	123	0.14	0.001
Grocery					
Cereal	0.09	-72.6	833	0.04	0.000
Coffee	0.37	-94.1	255	0.06	0.000
Crackers	0.13	-16.2	123	0.13	0.000
Snacks	0.76	-40.7	53	0.22	0.003
Soft drinks	1.03	-71.3	69	0.17	0.001
Water	0.12	-33.5	281	0.19	0.000
Packaged meat					
Bacon	0.08	-29.2	347	0.07	0.000
Dinner sausage	0.26	-48.3	187	0.19	0.000
Taxable grocery					
Bleach	0.04	-26.3	661	0.01	0.000
Laundry detergent	0.10	-78.0	805	0.03	0.000
Pet food	1.49	-38.2	26	0.42	0.005

Notes: Changes are relative to the myopic seller. Changes in first-period and second-period profit are in percent of uniform profit, and the change in two-period profit is in percent of the myopic seller's two-period profit. "Break-even" is the number of months needed to repay the cost of the best experiment. "Price deviation" is the largest deviation of the forward-looking seller's first-period prices from the uniform price, across the category's products.

cost still takes between 29 months (soft drinks) and 72 months (fresh eggs).

C.4 Incentive to withhold purchases

We evaluate one strategy by which a consumer could obtain lower personalized prices, making no purchase in a category during the 12-month learning phase. The seller’s pricing rule is that of our baseline, and we do not solve for an equilibrium. Let $CS_d(\mathbf{p})$ denote the expected monthly consumer surplus of a consumer of type d who faces the prices $\mathbf{p}$,

$$CS_d(\mathbf{p}) = -\frac{\rho_d}{\alpha_d} \ln \left(1 + \sum_{j=1}^J \exp(\delta_j + \beta_d + \alpha_d p_j) \right).$$

As in the counterfactuals, the environment is stationary, with the same product qualities and uniform prices in every month. The cost of the strategy is $12 CS_d(\mathbf{p}^U)$, the surplus of 12 months at uniform prices. A consumer with purchase history $\mathbf{h}_\ell$ faces the personalized prices $\mathbf{p}(\mathbf{h}_\ell)$. For each history $\mathbf{h}_\ell$, let $\mathbf{h}_\ell^0$ denote the history with the same shopping occasions and no purchase. A consumer who withholds purchases generates $\mathbf{h}_\ell^0$ instead of $\mathbf{h}_\ell$ and faces the prices $\mathbf{p}(\mathbf{h}_\ell^0)$. The monthly gain of a consumer of type d is the difference in surplus between these two price vectors, averaged over the histories of type d ,

$$G_d = \sum_{\ell=1}^L f_{\mathbf{h}_\ell|\Theta}(\mathbf{h}_\ell \mid \Theta_d) \left[CS_d(\mathbf{p}(\mathbf{h}_\ell^0)) - CS_d(\mathbf{p}(\mathbf{h}_\ell)) \right],$$

where $f_{\mathbf{h}_\ell|\Theta}$ is the probability of the history given the type, defined in equation (5). The break-even horizon, in months, is the ratio of the cost to the monthly gain,

$$\frac{12 CS_d(\mathbf{p}^U)}{G_d},$$

for the types with $G_d > 0$. We evaluate the sum in G_d by recursion, using the personalized prices of the counterfactual with 12-month histories.

Table C.6 reports, by category, the average extra discount obtained by making no purchase, the shortest break-even horizon across consumer types, and the share of consumers who break even within 12 months. We restrict our attention to the consumers who have something to gain from withholding. The extra discount is averaged over the consumers who would obtain a lower price, and the share is computed among the consumers who purchase in the category. The average extra discount lies between 0.8% (bleach) and 11.6% (soft drinks). The forgone surplus is large relative to the gain. In 23 categories, the shortest break-even horizon across

consumer types lies between 14.3 months (snacks) and 26.2 months (laundry detergent), so that no consumer breaks even within 12 months. Soft drinks is the exception. Its shortest horizon is 9.3 months, and two types break even within 12 months. They account for 6.9% of the consumers who purchase in the category.

Table C.6: Incentive to withhold purchases, by category

Category	Extra discount (%)	Shortest break-even (months)	Consumers breaking even within 12 months (%)
Alcohol			
Beer	3.3	19.2	0.0
Wine	7.8	17.2	0.0
Dairy			
Butter/margarine	4.1	20.1	0.0
Cheese	8.2	15.3	0.0
Creams	4.5	21.3	0.0
Fresh eggs	9.2	17.5	0.0
Refr. juices	8.2	15.5	0.0
Milk	5.2	15.4	0.0
Yogurt	7.2	18.0	0.0
Frozen food			
Ice cream	4.7	19.9	0.0
Fz. pizza	2.2	22.2	0.0
Fz. potatoes	3.9	22.0	0.0
Fz. vegetables	7.4	18.6	0.0
Grocery			
Cereal	5.1	20.1	0.0
Coffee	2.2	17.5	0.0
Crackers	7.5	16.9	0.0
Snacks	8.1	14.3	0.0
Soft drinks	11.6	9.3	6.9
Water	4.9	17.8	0.0
Packaged meat			
Bacon	7.0	20.2	0.0
Dinner sausage	5.5	22.6	0.0
Taxable grocery			
Bleach	0.8	19.5	0.0
Laundry detergent	4.7	26.2	0.0
Pet food	9.3	18.9	0.0

Notes: 12-month histories. “Extra discount” is the average reduction in price obtained by making no purchase, in percent of the uniform price, among the consumers who would obtain a lower price by doing so. Prices are unweighted averages across the category’s products. “Shortest break-even” is the minimum of the break-even horizon across consumer types. The last column reports, among the consumers who make at least one purchase in the category over the 12 months, the share whose type breaks even within 12 months.

The exercise has three limitations. It considers a consumer who withholds all purchases, whereas a consumer who forgoes only some purchases bears a lower cost for a smaller discount.

It holds shopping occasions fixed, whereas a consumer could also stay away from the store, at the cost of forgoing purchases in every category. It also holds the seller's pricing rule fixed, whereas a seller facing such consumers would account for their incentives. See [Villas-Boas \(2004\)](#) for the equilibrium of a monopolist facing forward-looking consumers.

C.5 Solving for optimal transfers in loyalty programs

Consumers know their own preferences. Let $d_i = (d_i^1, \dots, d_i^{24})$ collect consumer i 's types in the 24 categories, and let $\mathbf{h}_i^k$ be the consumer's 12-month purchase history in category k . Consumer i enrolls if

$$\sum_{k=1}^{24} CS_{d_i^k}^k(\mathbf{p}^k(\mathbf{h}_i^k)) + b \geq \sum_{k=1}^{24} CS_{d_i^k}^k(\mathbf{p}^{NP,k}),$$

where $CS_d^k(\mathbf{p})$ is the expected monthly surplus of a consumer of type d in category k at prices $\mathbf{p}$, $\mathbf{p}^k(\mathbf{h}_i^k)$ are the personalized prices of equation (7), $\mathbf{p}^{NP,k}$ are the uniform prices for non-participants in the loyalty program, and b is the monthly transfer. Write $e_i(d_i) \in \{0, 1\}$ for this decision.

We do not observe d_i , so we evaluate the enrollment decision under the posterior distribution of the consumer's types given their histories. The decision depends on the types in all categories jointly, so this expectation runs over every combination of the categories' types, which cannot be enumerated. We approximate it by draws. A consumer's types are linked across categories by the shopping occasion probability, which is common to all categories. We therefore draw the consumer's shopping occasion probability once, from its posterior, and the category tastes within each category from that category's posterior conditional on it. We use $R = 300$ draws per consumer. Consumer i 's probability of enrolling is the share of draws that enroll,

$$e_i = \frac{1}{R} \sum_{r=1}^R e_i(d_{ir}).$$

The seller's profit is

$$\Pi(b) = \sum_i \frac{1}{R} \sum_{r=1}^R \left[e_i(d_{ir}) \left(\sum_{k=1}^{24} \pi_{d_{ir}^k}^k(\mathbf{p}^k(\mathbf{h}_i^k)) - b \right) + (1 - e_i(d_{ir})) \sum_{k=1}^{24} \pi_{d_{ir}^k}^k(\mathbf{p}^{NP,k}) \right],$$

where $\pi_d^k(\mathbf{p})$ is the expected monthly profit from a consumer of type d in category k at prices $\mathbf{p}$.

The two scenarios differ in how the seller sets the non-enrollment uniform price, $\mathbf{p}^{NP}$. In the first, the seller does not update its beliefs about non-enrollees, so $\mathbf{p}^{NP} = \mathbf{p}^U$. In the second, the seller forms the distribution of types among non-enrollees from the types that opt out,

$$\phi_{d|NP}^k = \frac{\sum_i \sum_{r=1}^R (1 - e_i(d_{ir})) \mathbb{1}(d_{ir}^k = d)}{\sum_i \sum_{r=1}^R (1 - e_i(d_{ir}))},$$

and $\mathbf{p}^{NP,k}$ is the uniform price of category k that maximizes profit given $\phi_{d|NP}^k$. Enrollment decisions and $\mathbf{p}^{NP}$ are determined jointly, and we require $\mathbf{p}^{NP}$ to be a best response to the consumers who opt out at that price, which rules out equilibria in which prohibitive prices that no consumer faces induce full enrollment.

We solve for the optimal transfer by grid search over 121 monthly transfers, from a fee of \$1 to a reward of \$0.20 in steps of one cent. The personalized prices $\mathbf{p}^k(\mathbf{h}_i^k)$ and the surplus and profit of each draw at these prices do not depend on the transfer and are computed once. For each transfer, we compute the enrollment decisions and the seller's profit. In the first scenario, no iteration is needed. In the second, we compute the enrollment decisions and $\mathbf{p}^{NP}$ by fixed-point iteration:

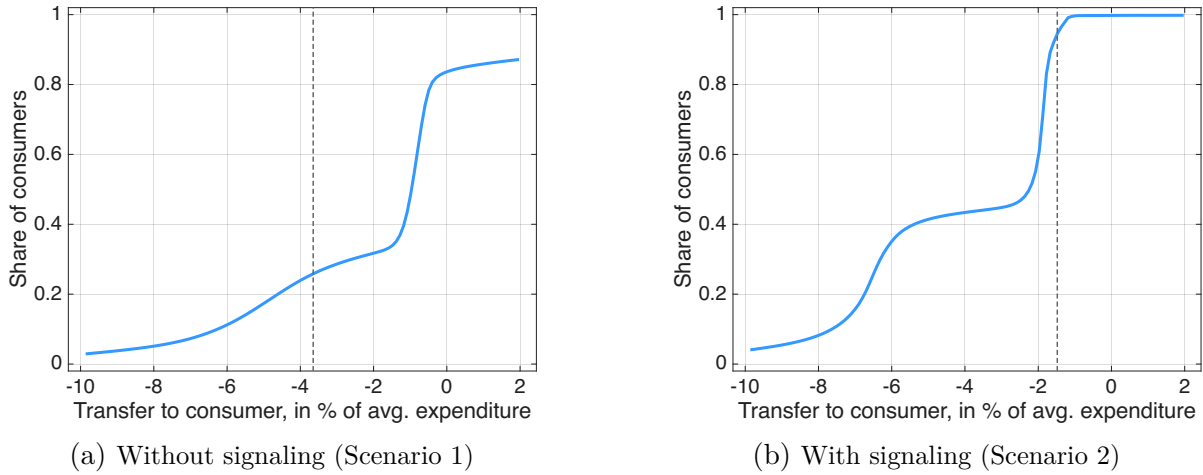
1. Start from $\mathbf{p}^{NP,k} = \mathbf{p}^{U,k}$ in every category and compute the enrollment decision of every draw.
2. Compute the distribution of types among the draws that do not enroll, $\phi_{d|NP}^k$, in every category.
3. Re-solve the uniform prices $\mathbf{p}^{NP,k}$ given $\phi_{d|NP}^k$.
4. Recompute the enrollment decisions at the new prices, and iterate steps 2 to 4 until no draw changes decision.

Two practical issues remain in computing this equilibrium. First, several equilibria may exist for a given monthly transfer, and the one reached depends on the enrollment decisions from which the iteration starts. We report the equilibria reached from no enrollment. We also start the iteration from four other enrollment sets at twelve transfers around the optimal fee. All starts reach the same equilibrium, except at fees of 1.78% to 2.27% of average expenditure, where the start with near-full enrollment reaches a second equilibrium with enrollment higher by 1–3 percentage points and profits higher by 0.2% to 0.4%.

Second, the iteration may cycle instead of converging once most consumers enroll: the non-enrollee price is then set on a small pool of types, a few of them change their decision when it moves, and their decisions move it back. We let the iteration run until the number of switching draws stops falling, then continue for 40 iterations and report the averages of profits, enrollment, and surplus over these iterations. The cycle is narrow: at the optimal fee, profits over the cycle stay within 0.2% of their average.

Figure C.6 shows how enrollment varies with the transfer. In Scenario 1 (panel a), the consumers who gain from personalized pricing enroll without any transfer. Some consumers are ready to pay substantial fees to access lower personalized prices: 17.1% still enroll at a monthly fee of 5% of average expenditure. Meanwhile, consumers who expect personalized prices above uniform levels must be highly compensated: even at a monthly reward of 2% of average expenditure, 12.8% refuse to enroll. In Scenario 2 (panel b), enrollment is higher at every transfer, since opting out exposes a consumer to the uniform prices set for the non-enrollee pool: 41.3% of consumers enroll at the 5% fee, against 17.1% in Scenario 1. Enrollment rises steeply just above the optimal fee, and exceeds 99% at every fee below 1.2% and at every reward.

Figure C.6: Loyalty program enrollment by transfer



Notes: Share of consumers who enroll, by monthly transfer expressed in percent of average consumer expenditure (\$10.1 per month). The vertical dashed lines indicate the optimal transfers (-3.65% in Scenario 1, -1.5% in Scenario 2).