Intertemporal Price Discrimination in Storable Goods Markets

By Igal Hendel and Aviv Nevo

We study intertemporal price discrimination when consumers can store for future consumption needs. We offer a simple model of demand dynamics, which we estimate using market-level data. Optimal pricing involves temporary price reductions that enable sellers to discriminate between price sensitive consumers, who stockpile for future consumption, and less price-sensitive consumers, who do not stockpile. We empirically quantify the impact of intertemporal price discrimination on profits and welfare. We find that sales (i) capture 25–30 percent of the gap between non-discriminatory profits and (unattainable) third-degree price discrimination profits, (ii) increase total welfare, and (iii) have a modest impact on consumer welfare. (JEL D11, D12, L11, L12, L81)

Consumers are heterogeneous in many ways, including preferences, income, and transportation and storage costs. Faced with heterogeneous consumers, firms can benefit from price discrimination. When consumer types are unobservable, firms need to rely on screening mechanisms to separate them. Empirically, we know little about the potential benefits of price discrimination, or how well different screening mechanisms work. Furthermore, the impact of price discrimination on welfare, especially in an oligopoly setting, is ambiguous.

In this paper, we empirically study the role of intertemporal price discrimination in storable goods markets. We estimate preferences and use the estimates to test whether price discrimination motives are driving sales. Finding that sales could be driven by price discrimination motives, we evaluate the effectiveness of sales as a price discrimination tool. We do so by comparing profits when there are sales to profits under third-degree price discrimination, where the seller can identify the different consumer types and prevent arbitrage.

In order to compute optimal pricing and welfare under different counterfactual pricing regimes, we need to model how consumers respond to a temporary price
reduction. Since the good is storable, buyers solve an inventory problem. To ensure that the problem is tractable we develop a simple dynamic demand model, which is easy to estimate using market-level data.

A key to the simplicity of the demand model is the storage technology: consumers can store for a pre-specified number of periods at no cost. To see how this assumption simplifies the problem, consider the simplest situation in which there is a single variety of the product, consumers can only store for one period, and they have perfect foresight (of future prices and demand needs). Because consumers can store for one period, period \( t \) consumption can be purchased in period \( t - 1 \) or in period \( t \). Absent discounting and storage costs, consumers will purchase in the period with the lowest price. The lesser of the prices in periods \( t \) and \( t - 1 \), which we call the effective price, is the relevant price for period \( t \) consumption. The consumers’ problem is equivalent to a static one, but replacing the actual prices with the effective ones. There is no need to solve a Bellman equation. This logic can be extended to multiple products, rational expectations about future prices, and longer storage periods.

The parameters of the model are identified from aggregate (store-level) quantities and observed prices. The intuition for identification is as follows. Consider the example above, in which consumers can store for at most one period. Holding period \( t \) price constant, we observe the quantity purchased at \( t \) if there was a sale in \( t - 1 \) and if there was no sale. We define below precisely what we mean by a sale, but for now it is a period where consumers should store—in this example, a period when prices are lower than in the following period. Absent dynamics, these two quantities should be the same on average. If they are not, dynamics are likely at play. Furthermore, by looking at how many lags impact current demand we can pin down the nature of the dynamics, such as the storage length.

We allow for different consumer types: storers and non-storers. Using store-level scanner data for soft drinks we find that consumers who store are more price-sensitive than consumers who do not. Heterogeneity in price sensitivity generates benefits from discrimination by targeting storers with lower prices. To evaluate profits and consumer surplus under different pricing regimes we use the estimated demand and the corresponding first-order conditions. We consider two benchmarks: non-discriminatory pricing and third-degree price discrimination under the assumption of no arbitrage. The third-degree discrimination benchmark is not feasible because, in practice, sellers cannot target storers and non-storers with different prices. We find that third-degree price discrimination would increase profits by 9–14 percent relative to non-discriminatory prices. Temporary price reductions, as a form of partial discrimination, enable sellers to capture around 24–30 percent of the potential gains generated by discrimination.

The welfare implications of third-degree price discrimination by a monopolist were studied by Robinson (1933), and later formalized by Schmalensee (1981), Varian (1985), and Aguirre, Cowan, and Vickers (2010) among others. The impact of discrimination on welfare is ambiguous. In oligopoly situations there are virtually no predictions as to how discrimination impacts welfare. We find empirically that total welfare increases under third-degree price discrimination as well as under intertemporal price discrimination. Sellers and consumers who store are better off. Consumers who do not store are worse off, but in most cases their loss is more than offset by storers’ welfare gains.
The demand estimates suggest that dynamics are important. Own-price responses from models that neglect dynamics are likely to be overstated (Hendel and Nevo 2006b). Static estimates reflect a weighted average of long-run price responsiveness (dictated by the underlying preferences) and short-run inventory (intertemporal) considerations. Biases in cross-price elasticities have been trickier to sign. Our model suggests that neglecting dynamics causes a downward bias in estimated cross-price effects. The reason is that standard static estimation controls for the wrong price of competing goods. Demand estimates can be used in antitrust, for example, to assess unilateral effects in merger analysis. Both biases, the upward bias in own-price effect and downward bias in cross-price effect, attenuate the computed unilateral effect.

Our estimates show that sellers have an incentive to intertemporally price discriminate, suggesting that sales could be motivated by price discrimination. Incentives for sales are similar to those of Jeuland and Narasimhan (1985) and Hong, McAfee, and Nayyar (2002) but in a somewhat different context. Hong, McAfee, and Nayyar (2002) model a homogeneous good sold in different stores, under unit demand, and single unit storage. Since our focus is primarily empirical we need a framework suited for demand estimation, one that allows for product differentiation and endogenous consumption and storage levels.

The model relates to Erdem, Imai, and Keane (2003) and Hendel and Nevo (2006b) who estimate structural models of consumer inventory behavior. Our demand model is motivated by this literature but offers substantial computational savings and a tractable supply side. Our results relate to several papers in the empirical price discrimination literature. For example, Villas-Boas (2009) uses demand estimates and pricing (in the vertical chain) to predict that banning wholesale price discrimination in the German coffee market would increase welfare. Lazarev (2012) looks at the welfare effects of intertemporal price discrimination by airlines.

**I. Motivating Facts**

**Figure 1** shows the price of a two-liter bottle of Coke in a store over a year. The pattern is typical of pricing observed in scanner data: regular prices with temporary price reductions, followed by a return to the regular price. With the exception of a short transition period, in any given 2–3 week window there are two relevant prices: a sale price and a non-sale price. We note that while sales are not perfectly predictable, they are quite frequent. Both of these facts will feed into our modeling below.

The pattern in Figure 1 raises two questions. How do consumers faced with this price process behave? What is the supply model that generates this pattern?

Since soft-drinks are storable, price patterns like those seen in Figure 1 create incentives to buy during a sale for future consumption. Indeed several papers in the economics and marketing literature document demand dynamics (see Blattberg and Neslin 1990 for a survey of the marketing literature). Boizot, Robin, and Visser (2001) and Pesendorfer (2002) show that demand increases in the duration from the previous time the product was on sale. Hendel and Nevo (2006a) document that duration from the previous purchase is shorter when purchasing during a sale, while duration until the following purchase is longer when purchasing on sale.
To see evidence of dynamics, we display in Table 1 the average quantity of two-liter bottles of Coke sold during sale and non-sale periods (we present the data in more detail later). During sales the quantity sold is substantially higher (623 versus 227, or 2.75 times more). More importantly, the quantity sold is lower if a sale was held in the previous week (399 versus 465, or 15 percent lower). The impact of previous sales is even larger if we condition on whether or not there is a sale in the current period (532 versus 763, or 30 percent lower if there is a sale and 199 versus 248, or 20 percent lower in non-sale periods).

We interpret the simple patterns present in Table 1 as evidence of demand dynamics: quantity purchased is correlated with previous prices. Moreover, Table 1 hints that storing behavior might be heterogeneous. If every consumer stored during sales we would expect quantity sold at a non-sale price that follows a sale to be quite small. The first row of the table shows that this is not the case.

We now turn to the second question: What supply model generates such a pricing pattern? The literature has offered several explanations. First, prices could be driven by changes in production costs or demand shocks. In practice, however, it is hard to imagine changes in static conditions that would generate this pricing pattern.

An alternative explanation focuses on consumer search behavior. Varian (1980) and Salop and Stiglitz (1982) offer models where firm pricing involves mixed strategies in equilibrium. While these models have some attractive features, they do not explain the regular sales patterns in Figure 1 and cannot explain the quantity patterns in Table 1.

\[1\] The patterns in Table 1 persist if we condition on whether or not Pepsi is on sale. Therefore, the patterns we see are not driven by Pepsi price differences across states.
A third theory (Lal 1990) assumes that some consumers are loyal to national brands and have a higher willingness to pay, while others will switch between national and generic brands for significant enough price differences. Lal (1990) shows that for certain parameter values, the national brands tacitly collude and take turns at pricing high, selling only to the loyal consumers, and low, selling to the switchers as well as their loyal consumers. While this model can explain the pricing patterns, it is inconsistent with the dynamic purchasing patterns we see in Table 1.

A fourth theory explains sales as part of retailer behavior and multi-category pricing (Lal and Matutes 1994). In these models price reductions are seen as loss leaders and meant to draw consumers into the store.

A final set of theories focuses on sales as intertemporal price discrimination. Sobel (1984); Conlisk, Gerstner, and Sobel (1984); and Pesendorfer (2002) present models of intertemporal price discrimination without repeat purchase (more recently used by Chevalier and Kashyap 2011), while Jeuland and Narasimhan (1985) and Hong, McAfee, and Nayyar (2002) do so for storable goods. We focus on this class of models and test whether heterogeneity in consumer preferences and storing ability can explain sales in our data. Finding that it can, we quantify the impact of sales.

II. The Model

In order to evaluate the impact of intertemporal price discrimination we need to compute equilibrium prices. Equilibrium prices are determined by the interaction between the sellers’ profit-maximization problem and the consumers’ inventory problem.

In its general form the problem is complex. Pricing, in principle, may depend on the inventory held by each of the buyers in the population as well as their preference shocks. However, consumption, and thus inventories, are private information. Sellers may use past purchases, or total quantity sold, to infer inventory holdings, but even such limited information is likely unavailable to buyers and competitors. It is therefore unclear on what information pricing strategies depend.

Hong, McAfee, and Nayyar (2002) provide a model in which past prices, which are public information, are sufficient to infer past purchases. In their model, preferences are deterministic and homogeneous, so that the starting inventory and the identity of those holding inventory is known at any point in time. With estimation in mind, we would like a more general model than Hong, McAfee, and Nayyar (2002), but one that still retains the sufficiency of public information. We want the model to allow for (i) differentiated products, (ii) consumption and the quantity stored to depend on price, (iii) non-deterministic demand, and (iv) heterogeneous consumers.
A. Demand

In this section we present a model that accommodates the empirical requirements, yet past prices are a sufficient statistic for the state.

1. The Setup: We make the following assumptions on the nature of consumer heterogeneity, storage costs, and consumer expectations.

ASSUMPTION 1: There are two types of consumers: storers and non-storers.

Assumption 1 implies that some buyers are unable, or unwilling, to store. One can view this as an assumption on the distribution of storage costs: some consumers have high storage costs and therefore will not find it profitable to store.

We allow storers and non-storers to have different quasi-linear preferences:

\[
U_t^S(q, m) = u_t^S(q) + m \quad \text{and} \quad U_t^{NS}(q, m) = u_t^{NS}(q) + m,
\]

where \(q = [q_1, q_2, \ldots, q_J]\) is the vector of quantities consumed of the \(J\) varieties of the product and \(m\) is the quantity consumed of a numeraire. We index preferences by \(t\) to allow for changing needs. Preferences could reflect a representative consumer, assuming the household level preferences are of Gorman form and yield an aggregate consumer. Alternatively, preferences could represent explicit aggregation over individual households (for example, as in Berry, Levinsohn, and Pakes 1995).

Allowing for preference heterogeneity is critical for our supply side analysis. Heterogeneity drives sellers’ incentives to discriminate. It is an empirical matter whether and how the preferences of the two consumer types differ. We may find that the proportion of either type is zero. Alternatively, we could find non-storers to be more price-sensitive. Either of these findings imply that discrimination is not motivating sales.

Absent storage, purchases equal consumption, thus, consumers’ aggregate demand by type is

\[
q_t^S = Q_t^S(p) \quad \text{and} \quad q_t^{NS} = Q_t^{NS}(p),
\]

where \(Q_t(\cdot)\) is the demand function implied by maximizing the utility in (1) subject to a budget constraint.

With storage, purchases and consumption need not coincide. In order to predict storers’ purchases we make the following assumptions.

ASSUMPTION 2 (Storage Technology): Storage is free, but inventory lasts for only \(T\) periods (fully depreciates afterwards).

Taken literally the assumption captures perishability. For example, \(T = 1\) fits products that last no more than two weeks (the period of purchase and the following one). An alternative interpretation of the assumption is contemplation; while the product may last longer, buyers only consider purchasing for \(T\) periods ahead when at the store. The assumption delivers simple dynamics: it simplifies the state and
detaches the storage decisions of the different varieties. Under A2, other products’
effective prices, which we will define below, are a sufficient statistic for quantity in
storage. Effective prices are public information and therefore this assumption not
only eases demand estimation but also helps formulate the sellers’ problem.

Finally, we assume the following about consumer expectations.

ASSUMPTION 3 (Future Demand Needs): Consumers know their future demand
needs at least $T$ periods ahead.

Since our focus is on low $T$, and time periods are weeks, this assumption seems
reasonable.

ASSUMPTION 4 (Perfect Foresight): Consumers have perfect foresight regarding
future prices.

Under perfect foresight the consumer problem becomes particularly simple. Obviously, perfect foresight is a strong assumption, especially if prices vary a lot. We also consider an alternative assumption:

ASSUMPTION 4’ (Rational Expectations): Consumers hold rational expectations
about future prices.

We present estimates under both Assumptions 4 and 4’ for $T = 1$; the results are
similar.

2. The Storing Consumer’s Problem: We first find storers’ demand under
Assumptions 1–4, and later consider Assumption 4’. Under perfect foresight, con-
sumers know in which of the $T$ preceding periods it is cheapest to purchase for
period $t$ consumption. We define the effective price as the minimum price in the re-
levant $T + 1$ periods: period $t$ and the $T$ preceding periods. The consumer’s problem
is then just like a static problem, but replacing the actual prices with the effective
ones. There is no need to solve a Bellman equation.

For ease of exposition we ignore discounting. The application involves weekly
data, and therefore discounting is not critical. In order to ensure a solution we focus
on a finite, $R$ period, horizon.

Storers, who start with no inventory, maximize the sum of utility over the
$R$ periods subject to a budget constraint and storing constraints:

$$\text{Max } \sum_{t=1}^{R} \mathbb{E}(u_t^S(q_t) + m_t) \text{ subject to }$$

$$\sum_{t=1}^{R} (p_t' x_t + m_t) \leq \sum_{t=1}^{R} y_t \quad \text{and} \quad q_t \leq x_t + \sum_{\tau=1}^{t-1} (x_\tau - q_\tau - e_\tau),$$

where $x_t$ is the vector of purchases and $e_\tau$ is the vector of unused units that expire
in period $\tau$. We have in mind items on which expenses are small relative to wealth,
$\sum_{t=1}^{R} y_t$, so that $m_t > 0$ for all $t$. Therefore, income effects or liquidity constraints in
any particular period do not play a role.
Define the effective price of product $j$ in period $t$ as:

$$p_{jt}^{ef} = \min\{p_{jT}, \ldots, p_{jt}\}.$$  

(3)

Notice that the effective price is actually the opportunity cost, since products are not purchased at any price other than the effective price. Denoting by $p_{t}^{ef}$ the vector of period $t$ effective prices, the problem of the storing consumer becomes

$$\max \sum_{t=1}^{R} \mathbb{E}(u_{t}(q_{t}) + m_{t}) \text{ subject to } (p_{t}^{ef} q_{t} + m_{t}) \leq y_{t}.$$  

The optimization of the storer is equivalent to $R$ static optimization problems where, in each of them, prices are $p_{t}^{ef}$. Optimal consumption in period $t$ is $q_{t}^{*} = Q_{t}^{*}(p_{t}^{ef})$. It is not necessary to solve a dynamic problem since, for each period $t$, the consumer solves an independent static problem. Period $t$ consumption is purchased at the lowest of the $T + 1$ prices. The dynamics are addressed simply by replacing prices with effective prices.

The storage of a product affects the purchases and consumption of all other products exclusively through its effective price. To see that, suppose that product $j$ was purchased during a sale for later consumption. The demand for all other products is naturally affected by the fact that product $j$ is in the pantry waiting to be consumed. Since $p_{jt}^{ef}$ was known at the time that all other varieties to be consumed at period $t$ were purchased, the impact of $j$ inventory on demand for $-j$ is fully captured by $p_{jt}^{ef}$.

3. Purchasing Patterns: Consumption is the solution to a sequence of static problems where prices are replaced by effective prices. Characterizing predicted purchases requires a bit of accounting and some unpleasant notation (to take care of tie-breaks in the timing of purchases).

Purchases by storers in period $t$ are the sum over current and future consumption needs supplied by current purchases. More precisely, period $t$ purchases of product $j$ are computed by first comparing $p_{jt}$ to the $T$ preceding prices; if $p_{jt}$ is the lowest, the consumer is predicted to purchase today for current consumption. Next we compare $p_{jt}$ to the vector $[p_{jtT+1}, \ldots, p_{jt+T}]$ to see if the consumer buys at $t$ for $t+1$ consumption. We repeat this up to period $t+T$. Thus, $T + 1$ price comparisons tell us for which of the coming periods the consumer is purchasing today. For each of those purchases, quantity purchased is determined by the static demand in equation (2) using the current price of $j$, $p_{jt}$, and the effective price of other products $p_{jt+\tau}$, where $\tau = 0, \ldots, T$. Period $t$ purchases are thus

$$X_{jt}^{S}(p_{t-T}, \ldots, p_{t+T}) = \sum_{\tau=0}^{T} Q_{jt+\tau}(p_{jt}, p_{jt+\tau}) \mathbb{I}[p_{jt} = p_{jt+\tau}].$$

(4)

When there are ties (caused by the effective price being available more than once) we assume that buyers purchase immediately if the price is below a threshold,

---

2 We assume $p_{jT} = \infty$ for $T > t$, namely, purchases before period 0 are not feasible.
prices. As a practical matter we use the median price as the threshold.

4. Predicted Behavior for $T = 1$: We now spell out what equation (4) implies in the case of $T = 1$. Focusing on $T = 1$ serves two purposes. First, it clarifies the model and aids in the discussion of identification in the next section. Second, it prepares the ground to show how the model works under rational expectations.

Equation (4) determines whether storers purchase in period $t$ for consumption in periods $t$ to $t + T$. For $T = 1$ purchasing behavior is determined by whether the period $t$ consumption is already in storage, and whether the consumption for period $t + 1$ is purchased at period $t$ or at period $t + 1$. Equation (4) can be presented in a more convenient form. Define sale periods, $S_t$ for a specific product $j$ as those periods in which the storing consumer buys for future consumption, namely if $p_{j,t} < p_{j,t+1}$. A non-sale period, $N_t$, is one in which $p_{j,t} > p_{j,t+1}$.

Assume for a moment that only product $j$ can be stored. Thus, to predict storers’ behavior we only need to define four events (or types of periods): a sale preceded by a sale (SS), a sale preceded by a non-sale (NS), a non-sale preceded by a sale (SN), and two non-sale periods (NN). Given Assumptions A1–A4, storers’ purchases at time $t$ are given by

\[
x_{j,t}^S(p_{t-1}, p_t, p_{t+1}) = \begin{cases} 
Q_{j,t}^S(p_{j,t}, p_{j,t}) & 0 \quad \text{NN} \\
0 & 0 \quad \text{SN} \\
Q_{j,t}^S(p_{j,t}, p_{j,t}) + & Q_{j+1,t}^S(p_{j,t}, p_{j,t+1}) \quad \text{NS} \\
0 & Q_{j+1,t}^S(p_{j,t}, p_{j,t+1}) \quad \text{SS}
\end{cases}
\]

where $Q_{j,t}^S()$ is the static demand of storers (defined in (2)).

At high prices there is no incentive to store, in which case purchases equal consumption, given by $Q_{j,t}^S()$, if there was no sale last period, or zero if there was a sale in the previous period (i.e., in $SN$ consumption is out of storage). During sales preceded by a non-sale period, purchases are for current consumption as well as for inventory. During periods of sale preceded by a sale, current consumption comes from stored units, so purchases are for future consumption only, $Q_{j+1,t}^S(p_{j,t}, p_{j,t+1})$. Notice the difference in the second argument of the anticipated purchases, $Q_{j+1,t}^S(p_{j,t}, p_{j,t+1})$, relative to purchases for current consumption, $Q_{j,t}^S(p_{j,t}, p_{j,t})$. Purchases for future consumption take into account the future price, and therefore future consumption, of products $-j$.

When all products are storable, the way to account for storability is to replace $p_{j,t+1}$ by $p_{j,t+1}^m$, as discussed in the previous section. For example, consider the

---

3 If $p_{j,t} = p_{j,t+1}$ then the period is classified as a sale if $p_{j,t}$ is below a threshold $p_j^m$ and a non-sale otherwise.
event \( NN \) (product \( j \) is not on sale at \( t \) or at \( t - 1 \)) and assume that product \(-j\) was on sale at \( t - 1 \) (but is not on sale at \( t \)). Storers’ purchases are \( Q^S(j; p_{jt}, p^S_{-jt}) \) instead of \( Q^S(j; p_{jt}, p^S_{-jt}) \).

There are two important implications of equation (5) for estimation. First, equation (5) dictates in which states to scale demand up or down to account for storage and consumption out of storage.

Second, we can see from equation (5) that contemporaneous prices of other products are often the wrong control in the estimation. Using the current (cross) price \( p_{-j} \) generates a bias in the estimated cross-price effect. To understand the source of the bias consider what happens to purchases of \( j \) as a substitute’s sale ends. During the \(-j\) sale period naturally \( j \)'s demand goes down, but the increase in \( p_{-j} \) after the sale is not accompanied by an increase in the demand for \( j \), since the effective price of \(-j\) is still the sale price. Static estimates will misinterpret the lack of decline in \( j \)'s demand as \( p_{-j} \) increases (when the sale ends) as lack of price sensitivity, while in practice the effective price has not changed.

Allowing for higher \( T \) in the estimation is immediate. Equation (5) requires adjustment to reflect that consumers can plan more periods ahead, as reflected in (4).

5. Rational Expectations: We now explore demand under Assumption 4', rational expectations of future prices, while maintaining Assumptions 1–3 and \( T = 1 \). Equation (5) is still valid in the single product case (i.e., if other products are either not present or are sold at a constant price) but we need an appropriate classification of \( N \) and \( S \) periods. Naturally, we cannot define a sale based on \( t + 1 \) prices, which (absent perfect foresight) are not yet known.

Given the distribution of future prices, as shown in Hong, McAfee, and Nayyar (2002) and Perrone (2010), there is a price threshold below which the good is purchased for future consumption. Above the threshold buyers do not store. Below the threshold consumers have an incentive to store, even if some sales are deeper than others. We denote these periods as sales, \( S \), and non-storing periods as \( N \). Thus, in the single-product case the only modification, relative to the previous analysis, is in how we define a sale.\(^3\)

If more than one product is storable, then at the time that product \( j \) is purchased, the (effective) prices of other products may not yet be known. They would be known, for example, if the other products are on sale, in which case the effective price is the current price. In general, however, when some other product is not currently on sale, the consumer has to purchase product \( j \) under uncertain future \(-j\) prices. In the analysis below, we deal with this problem by only using a subset of the data to estimate the model.

Notice that predicted behavior requires knowledge of future prices only in some states. For example, if both products are in state \( SS \), demand does not depend on

---

\(^4\) Absent additional assumptions on the stochastic evolution of prices, the threshold is state specific. However, with some further structure (for example, that prices are i.i.d. or that future prices only depend on whether there is a sale at \( t \) (but not \( p_t \)), then there is a unique threshold.

\(^5\) The only minor subtlety is whether consumers who stockpiled buy additional units when the realized price is lower than the preceding price. One can assume that consumers purchase additional units at the low price, or that they pay no attention to products in categories for which they already stored (they do not even go through that aisle). Equation (5), as written, assumes the latter.
future prices. In states where demand does not depend on future prices, the predicted behavior as explained in the previous section is still valid, and can be used in the estimation under rational expectations. It turns out that for $J = 2$, in 10 of the 16 states (composed of the cross product of the four states of each good), predicted purchases of each product are independent of future prices. In total, only two out of the 16 states involve behavior that differs for both products. We elaborate below.

We also offer an alternative approach that uses all the data, but is more computationally intense. We describe this approach in the Appendix.

B. Seller Behavior

To compare pricing regimes we need to compute optimal seller behavior. We start by considering a single-product monopolist who can commit to future prices; later we argue that the same predictions arise absent commitment, and under duopoly with commitment. The monopolist faces the population of storers and non-storers described in the previous section, with flow demands $Q^S_t(p)$ and $Q^N_t(p)$, respectively. We will assume $Q^S_t(p) = Q^S(p) + \varepsilon^S_t$ and $Q^N_t(p) = Q^N(p) + \varepsilon^N_t$, where $\varepsilon^S_t$ and $\varepsilon^N_t$ are i.i.d. shocks with zero mean. These shocks reflect varying needs.

We assume that firms do not observe these shocks when setting prices and therefore optimal prices do not depend on them. Furthermore, because the shocks are i.i.d., knowing lagged shocks does not generate dynamics in pricing. Thus, in what follows, we drop the time subscript. We assume that marginal cost, denoted by $c$, is constant over time. Below we consider the implications of (stochastically) varying costs. We assume throughout the concavity of the various objective functions.

Denote by $p^*_S$ and $p^*_N$ the prices that maximize static expected profits from selling to the populations of storers and non-storers separately. Also let $p^*_N$ denote the optimal price of a non-discriminating monopolist facing the whole population. In other words, $p^*_N$ maximizes $(Q^N(p) + Q^S(p))(p - c)$. Denote $\pi^*_N = (Q^N(p^*_N) + Q^S(p^*_N))(p^*_N - c)$.

**Timing and Strategies.**—At time 0 the monopolist commits to the sequence of prices $\{p_1, \ldots, p_R\}$, where $R$ is the length of the horizon. As before, due to the lack of discounting, we assume that $R$ is finite to avoid an infinite sum. Storers, who under Assumption 4 know the price sequence, behave as described in (5). For $T = 1$ demand is a function of the prices at periods $t - 1, t$, and $t + 1: X_t(p_t|p_{t-1}, p_{t+1}) = Q^N_t(p_t) + X^S_t(p_t|p_{t-1}, p_{t+1})$. The monopolist maximizes

$$
\sum_{i=1}^{R} \pi(p_{i}|p_{i-1}, p_{i+1}) = \sum_{i=1}^{R} X_t(p_{i}|p_{i-1}, p_{i+1})(p_{i} - c).
$$

The maximization boils down to picking a sequence of prices that maximize profits, taking into account that if $p_{t-1} \leq p_t$, storers purchase for future use.

---

6 Commitment seems an adequate assumption for our application. Supermarkets set prices several weeks in advance.
1. Two-Period Problem: The two-period problem helps to highlight the gains from intertemporal price discrimination, and the conditions that make discrimination profitable.

If prices are constant or declining, no consumers store. In this case, optimal prices are \( p_{ND}^* \) in both periods and seller profits over the two periods are \( 2\pi_{ND}^* \).

If instead prices increase over time, then storers purchase in the first period for consumption in both periods. In the second period, because only non-storers are present, the optimal price is \( p_{NS}^* \). Increasing prices from the first to the second period allows the seller to partially separate consumers: in the second period only non-storers buy. Denote the intertemporal price-discriminating profits with first period (sale) price \( p \) as \( \pi_{IPD}(p) \). Profits are given by

\[
\pi_{IPD}(p) = (Q_{NS}^*(p) + 2Q_S(p))(p - c) + Q_{NS}^*(p_{NS}^*)(p_{NS}^* - c).
\]

The first term represents profits during the initial period, targeting storers who purchase for two periods, and non-storers for one period. The last term represents profits from non-storers in the second period.

Optimal first and second period prices, \( p \) and \( p_{NS}^* \), solve the following first order conditions:

\[
\begin{align*}
6 \quad & p_{NS}^* = c - \frac{Q_{NS}^*(p_{NS}^*)}{\frac{\partial Q_{NS}^*}{\partial p} \bigg|_{p=p_{NS}^*}} \\
 & p = c - \frac{Q_{NS}^*(p) + 2Q_S(p)}{\frac{\partial (Q_{NS}^*(p) + 2Q_S(p))}{\partial p} \bigg|_{p=p}}.
\end{align*}
\]

Increasing or constant prices could be optimal. For example, if \( p_{NS}^* \leq p_S^* \) then constant prices would do better. However, if

\[
7 \quad \pi_{IPD}(p) > 2\pi_{ND}^*
\]

for some \( p_S^* < p < p_{NS}^* \) then the pair \( \{p, p_{NS}^*\} \) does better than twice the non-discrimination profits. Whether constant or increasing prices are optimal depends, using the terminology of Robinson (1933), on whether non-storers are the strong market and storers are the weak market.

2. Multiple-Period Problem: With more than two periods, under condition (7) and commitment, optimal monopoly prices follow a cyclical pattern. A sale period is followed by \( T \) non-sale periods, during which only non-storers purchase. We show that price cycles of length \( T + 1 \) maximize the discriminating opportunities.

\[7 \quad \text{In this section we consider } R > T + 1, \text{ where } R \text{ is the length of the horizon. If } R \leq T + 1, \text{ the analysis of the previous section is still valid, with optimal prices being } p \text{ for one period, to supply storers, followed by } p_{NS}^* \text{ afterwards.} \]
The simplest case to consider is \( T = 1 \).

**Proposition 1:** Under condition \((7)\) optimal pricing involves cycles of \( p \) followed by \( \bar{p} \), with \( p^S < p < \bar{p} = p^NS \)

**Proof:**

See the Appendix.

It is interesting to notice that the cycle is designed to maximize the number of times that the seller discriminates. This prediction is similar to that of Jeuland and Narasimhan (1985).

Up to now we have assumed that the seller can commit to future prices. It is tedious but not difficult to show that the solution is time consistent. It involves showing that there is no profitable deviation for the monopolist in any subgame. To see why, absent commitment, the monopolist would price as in Proposition 1, consider a four period problem. In the last period, after a sale, \( p^NS \) is optimal. In the third period, the second cycle starts; it is easy to show that the low-high strategy is optimal. In the second period the seller could be tempted to induce storage, however, the continuation of the game after such a price involves lower profits, as it misses one of the price discrimination opportunities.

3. Duopoly: In the context of our application there is more than one manufacturer. We now argue that in equilibrium, under commitment, both sellers charge cyclical prices. The game involves two players, each simultaneously committing to a sequence of prices \( \{p_j, \ldots, p_{jR}\} \). As before, prices are set at time 0, prior to observing the time-varying demand needs. Storers behave as in (5), basing their period \( t \) consumption on the lowest of \( t \) and \( t - 1 \) prices, which by A4 are known in advance. We discuss the Nash equilibrium of this game.

Consider firm \( j \) taking as given firm \( -j \)'s cyclical price sequence. Storers’ demand is \( X^S_j(p_j, p_{-j}) \) because they always purchase good \( -j \) at the sale price. That is, demand is the same in every period regardless of the actual value of \( p_{-j} \). In contrast, the non-storers’ demand \( Q^NS_j(p_j, p_{-j}) \) depends on the contemporaneous \( -j \) price.

Define firm \( j \)'s non-discriminating profits given \( p_{-j} \) as \( \pi^*_{ND}(p_{-j}) \). Based on \( \pi^*_{ND}(p_{-j}) \) we can now modify condition (7) for the duopoly case. The following is a sufficient condition for discrimination to be profitable in a two-period problem

\[
(8) \quad (Q^NS_j(p_j, p_{-j}) + 2Q^S_j(p_j, p_{-j}))(p_j - c) + Q^NS_j(p_j, p_{-j})(\bar{p}_j - c) > \pi^*_{ND}(p_{-j}) + \pi^*_{ND}(\bar{p}_{-j}).
\]

The first term accounts for profits during the sale, when storers buy for both periods and non-storers for current consumption. The second term represents profits from non-storers. The right-hand side represents non-discriminating profits, the first term while product \( -j \) is on sale, and the second term for the competitor’s non-sale period. Condition (8) guarantees that in a two-period problem, firm \( j \), taking as given \( -j \)'s low-high price sequence, benefits from intertemporal price discrimination. Namely, \( j \) best-responds using cycling prices as well. Notice that
condition (8) is based on endogenous equilibrium values; it is meant to reflect fundamentals that deliver more profits under discrimination. The profitability of sales—namely, of discrimination—is an empirical matter, which we will evaluate with the estimated demand.

The same reasoning used to show that cycles are optimal for the monopolist in a longer horizon applies to the duopolist problem as well.8

III. Data and Estimation

A. Data

The dataset we use was collected by Nielsen and includes store-level weekly observations of prices and quantity sold for the 52 weeks of 2004 at 729 stores that belong to eight different chains throughout the Northeast United States. We focus on two-liter bottles of Coke, Pepsi, and store brands, which have a combined market share of over 95 percent.

There is substantial variation in prices over time and across chains. A full set of week dummy variables explains approximately 20 percent of the variation in the price in either Coke or Pepsi, while a full set of chain dummy variables explains less than 12 percent of the variation.9 On the other hand, a set of chain-week dummy variables explains roughly 80 percent of the variation in price, suggesting similarity in pricing across stores of the same chain (in a given week), and different pricing across chains (in a given week). It seems that all chains charge a single price in each store each week (i.e., there is no intra-week price variation). However, three of the chains appear to define the week differently than Nielsen. This results in a change in price mid-week, as defined in our data, which implies that in many weeks we observe a quantity weighted average price rather than the actual price charged. We therefore use the five chains for which we observe the prices charged.

Figure 2 displays the distribution of the price of Coke in the five chains we study below. For some of the estimation under rational expectations below we need to define sale prices—namely, prices (or price ranges) at which consumers stockpile. The distribution seems to have a break at a price of one dollar. We therefore define any price below a dollar as a sale. This is an arbitrary definition. A more flexible definition may allow for chain-specific thresholds, or perhaps thresholds that change over time. We prefer to err on the side of simplicity. Using this definition we find that approximately 30 percent of observations are defined as a sale for Coke and 36 for Pepsi.

For the analysis below we use 24,674 observations from five chains. The descriptive statistics for the key variables are presented in Table 2.9

---

8 There is an additional issue, the timing of the sales. It is possible that non-coincidental sales are more profitable than coincidental sales. The issue does not arise in a two-period setup since there is no point of having a sale in the second period. We abstract from the timing issue, which could arise in a longer horizon. A condition like (8) would determine optimal timing.

9 These statistics are based on the whole sample, while the numbers in Table 2 below are based on only five chains, as we explain next.
In this section we explain how we go from the model to estimation.

**Overview.**—We recover the utility parameters by matching weekly store-level purchases observed in the data to those predicted by the model. The observed purchases, $X_{jt}$, are

$$X_{jt} = Q_{j}^{NS}(p_t) + X_{j}^{S}(p_{t-T}, \ldots, p_{t+T}) = Q_{j}^{NS}(p_t) + X_{j}^{S}(p_{t-T}, \ldots, p_{t+T}) + \varepsilon_{jt},$$

where $Q_{j}^{NS}(p_t)$ are non-storers’ purchases defined by equation (2), $X_{j}^{S}(p_{t-T}, \ldots, p_{t+T})$ are storers’ purchases given by equation (4) (which in the case of $T = 1$ equals the formula given in equation (5) with effective prices replacing prices of the $-j$ goods),

---

**Figure 2: The Distribution of the Price of Coke**

*Note:* The figure presents a histogram of the distribution of the price of Coke over 52 weeks in 729 stores in our data.

**Table 2—Descriptive Statistics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Percent of variance explained by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chain</td>
</tr>
<tr>
<td>$Q_{Coke}$</td>
<td>446.2</td>
<td>553.2</td>
<td>5.6</td>
</tr>
<tr>
<td>$Q_{Pepsi}$</td>
<td>446.0</td>
<td>597.8</td>
<td>2.8</td>
</tr>
<tr>
<td>$P_{Coke}$</td>
<td>1.25</td>
<td>0.25</td>
<td>7.1</td>
</tr>
<tr>
<td>$P_{Pepsi}$</td>
<td>1.19</td>
<td>0.23</td>
<td>7.5</td>
</tr>
<tr>
<td>Coke sale</td>
<td>0.30</td>
<td>0.46</td>
<td>6.4</td>
</tr>
<tr>
<td>Pepsi sale</td>
<td>0.36</td>
<td>0.48</td>
<td>9.3</td>
</tr>
</tbody>
</table>

*Notes:* Based on 24,674 observations for five chains, as explained in the text. A sale is defined as any price below one dollar.
and $\varepsilon_{jt}$ is an error term discussed in the next subsection. To compute the predicted purchases we need to parametrize the demand functions $Q^{NS}_j$ and $Q^S_j$. In addition, we need to define the price regimes; that is, when storers purchase for future consumption or when they do not purchase at all. We explain both of these below.

Finally, for estimation we need to make an assumption on the error term. For simplicity, we assume $E(\varepsilon_{jt} | p_{t-T}, \ldots, p_{t+T}) = 0$ and therefore estimate the parameters of the demand functions using nonlinear least squares. If, however, prices are correlated with the error term then we would replace prices with instrumental variables in the conditioning set and use GMM to estimate the parameters.

From Model to Data.—We now discuss how the model and the empirical specification fit together.

First, consider the error term in the demand equation, which captures randomness in purchases conditioning on prices. We assume $\varepsilon_{jt}$ is additive, i.i.d., and unobserved by firms at the time of pricing. The inclusion of the shock is consistent with the theoretical demand model, where we allow consumers to have varying needs.

Second, we discuss the variation in prices and where this variation comes from. The prediction of the model, given in Proposition 1, is that price takes on two values (and that pricing follows a deterministic cyclical pattern). While the sale and non-sale pattern is apparent in the data, it is not deterministic and it typically involves more than two prices. The question is what explains the discrepancy. Note that demand shocks are unknown to firms at the time of pricing, and therefore cannot explain variation in observed prices.

A natural explanation for price variation is marginal cost variation. A more general version of Proposition 1, which allows for random costs, generates more than two prices as well as randomness in when sales are held (proof available upon request).

Finally, in the theoretical model, prices play two roles. First, prices determine how much is consumed. For this role, both in the theory and in the data, prices are used as a continuous variable. Second, prices determine the timing of purchases, namely, the state. During sales, storers purchase for future consumption. We classify sales, or price regimes, using the history of prices. Under perfect foresight, period $t$ is defined as a sale if $p_{jt} \leq p_{jt+1}$. Under rational expectations we use a price threshold to define sale periods. We use these definitions to determine whether storers purchase for future consumption (i.e., if a period is a sale period). But the prices that enter the various demand functions are the actual prices observed in the data.

Identification.—To see intuitively how preferences are identified, consider the case of $T = 1$. For simplicity assume a single product. All the arguments below go through with multiple products, but the conditioning vector needs to include effective prices of other goods. For any current price, $p_n$, we observe average purchases, $X_j(p_n | \text{sale at } t - 1) = \frac{1}{\#t} \sum_{j} X_{jt}(p_n | \text{sale at } t - 1)$ and $X_j(p_n | \text{no sale at } t - 1) = \frac{1}{\#t} \sum_{j} X_{jt}(p_n | \text{no sale at } t - 1)$, where a “sale” is defined above. The averaging is over time, namely, over realizations of $\varepsilon_{jt}$. The quantities observed for each $p_n$ can be used to recover the unknown quantities $Q^{NS}_j(p_n)$ and $Q^S_j(p_n)$ (which in turn identify the preferences of both groups).

To see how this works, we distinguish between two cases: when the current period is a sale and when it is not. We start with the case where $p_{jt}$ is not a sale price. In
this case if there was a sale at $t-1$ then storers do not purchase. Therefore, the observed purchase equals the demand of non-storers, $X_j(p_t | \text{sale at } t-1) = Q^\text{NS}_j(p_t)$. Furthermore, if there was no sale at $t-1$ then both storers and non-storers purchase only for current consumption; therefore, since $p_{jt}$ is not a sale price then $X_j(p_t | \text{no sale at } t-1) = Q^\text{NS}_j(p_t) + Q^\text{S}_j(p_t)$. Combining with the previous result, which recovers the demand of non-storers, we get $Q^\text{S}_j(p_t) = X_j(p_t | \text{no sale at } t-1) - X_j(p_t | \text{sale at } t-1)$. Thus, the demanded quantity for both storers and non-storers can be recovered from observed purchases.

We now turn to the case where $p_{jt}$ is a sale. In this case, if there was a sale at $t-1$ then $X_j(p_t | \text{sale at } t-1) = Q^\text{NS}_j(p_t) + Q^\text{S}_j(p_t)$ since non-storers purchase for current consumption and storers purchase for consumption at $t+1$. Similarly, $X_j(p_t | \text{no sale at } t-1) = Q^\text{NS}_j(p_t) + 2Q^\text{S}_j(p_t)$ since storing consumers purchase for consumption at $t$ and at $t+1$. This implies

$$Q^\text{S}_j(p_t) = X_j(p_t | \text{no sale at } t-1) - X_j(p_t | \text{sale at } t-1),$$

and

$$Q^\text{NS}_j(p_t) = 2 \times X_j(p_t | \text{sale at } t-1) - X_j(p_t | \text{no sale at } t-1).$$

So as before, we can recover the unobserved demanded quantities from the observed purchases.

Once we recover the demanded quantities from the observed purchases, the utility parameters can be identified following standard arguments. Note that in principle for some sale definition and price distributions there is a price range that is sometimes a sale and sometimes not.\(^{10}\) For these prices demand is actually overidentified.

For $T > 1$, we follow the same arguments but simply condition on a longer history of past events. The accounting becomes a bit more complex, but the basic idea is the same.

To further understand how the data separate different storing horizons, consider the following. If $T = 0$ (no consumers store) then $X_j(p_t | \text{sale at } t-1) = X_j(p_t | \text{no sale at } t-1)$ for all $p_t$. As we showed in the previous paragraphs, this equality does not hold under the $T = 1$ model. The differences across columns in Table 1 reflect such gaps, due to the impact of history on purchases. Intuitively, the static model says that current purchases are not a function of lagged prices, while the dynamic model allows for dependence. The same logic allows us to separate between the $T = 1$ model and $T = 2$ models: the $T = 1$ model says that lags beyond $t-1$ do not matter, while the $T = 2$ model says that these longer lags matter. By conditioning on further lags the data can distinguish between models.

With $T > 1$ the model is generally overidentified since we have many more relevant histories on which to condition. We can use these additional relevant histories to enrich the model and allow for more types. In the analysis below we allow for three types: non-storers ($T = 0$), storers with $T = 1$, and long horizon storers with $T = 2$ (or with $T = 3$).

\(^{10}\) For example, assume that $p_{jt} = p_{jt}$ for periods $t$ and $\tau$. Furthermore, assume that $p_{jt} \leq p_{j,t+1}$ and $p_{jt} > p_{j,t+1}$, then using the perfect foresight definition of a sale $p_{jt}$ is a sale price, while $p_{jt}$ is not.
**Estimation.**—For estimation we assume that demand for product $j$ at store $s$ in week $t$ is log-linear:

$$
\log q_{jst}^h = \omega^h \alpha_{sj} - \beta_j^h p_{jst} + \gamma_j^h p_{ist} + \varepsilon_{jst},
$$

$\quad j = 1, 2, \quad i = 3 - j, \quad h = S, NS,$

where $\alpha_{sj}$ is a store-specific intercept for each brand, and $\varepsilon_{jst}$ is an i.i.d. shock. The parameters $\omega^h$ allow for different intercepts for each consumer type. We scale these parameters to add to one and define $\omega = \omega^{NS} = 1 - \omega^S$. The parameter $\omega$ represents the fraction of non-storers when all prices are zero.

We also experimented with a linear demand specification. In general, the results, such as the difference between the static and dynamic model and the heterogeneity across storers and non-storers, are similar. However, in the linear model predicted demand can be negative (especially storers’ demand at a high price). The log-linear specification avoids the negativity problem.

We estimate all the parameters by non-linear least squares. We minimize the sum of squares of the difference between the observed and predicted purchases. In principle, we could use instrumental variables to allow for correlation between prices and the econometric error term. However, we do not think correlation between prices and the error term is a major concern in the application below. To obtain the exact estimating equations we combine equations (5) and (9) and allow for store fixed effects. To account for the store-level fixed effects we de-mean the data, which makes all the parameters enter the equations non-linearly. We show in the Appendix how to modify the estimating equation to account for the fixed effects.

**Estimation under Rational Expectations.**—As we discussed in Section IIA5, defining the purchases by storers is more complex under rational expectations. In particular, absent perfect foresight, defining effective prices is problematic. The approach we follow below involves estimating the model as in the perfect foresight case, but using only part of the sample. An alternative, discussed in the Appendix, uses all the data but involves solving a system of equations.

With two products and $T = 1$, there are 16 states (four for each product). In ten of these states the purchases predicted by the rational expectations model coincide with predicted purchases under the perfect foresight model, using the same definition of a sale. We can therefore recover preferences under rational expectations by restricting the sample to those states. The reason predictions coincide is that demand in those states does not depend on expected future prices. Thus, whether we assume perfect foresight or rational expectations is immaterial. For example, suppose the Coke state is $NN$. If the Pepsi state is either $NN$ or $NS$ then both models imply that Coke purchases depend on the current price of Pepsi, as the relevant cross price. However, if the Pepsi state is either $SS$ or $SN$ then the models differ in their predictions. Under both models the consumer bought Pepsi at $t - 1$ for consumption at $t$; however, the models differ in how much was purchased (since the future effective
Coke price is known in one case, and not in the other), and therefore, also differ in how much Coke is bought at $t$.\textsuperscript{11}

\section*{IV. Results}

\subsection*{A. Demand Estimates}

The estimation results are presented in Tables 3 and 4. The dependent variable is the (log of the) number of two-liter bottles of Coke or Pepsi sold in a week in a particular store. All regressions include store fixed effects as well as the price of the store brand.

1. Main Results: The first two columns in Table 3 display estimates from a static model. The rest of the columns present estimates from the dynamic model under different assumptions. In all cases we assume $T = 1$ and allow for different price sensitivity for storers and non-storers. We also impose two restrictions. First, we impose that the fractions of storers for Coke and Pepsi are the same. We could allow for two parameters, but consistent with the idea of a population of storers who decide what product to purchase, we impose the same parameter. Second, the cross-price effect between Coke and Pepsi is imposed to be symmetric.

In columns 3 and 4 we present estimates under perfect foresight (A4) and defining sales based on actual prices: consumers stockpile if prices at $t$ are lower than $t + 1$ prices. In the next set of columns, 5 and 6, we continue to assume perfect foresight but use a different definition of a sale. Now a sale is defined as any price below one dollar: whenever buyers observe a price below a dollar they purchase for future consumption.\textsuperscript{12,13} In both cases, the definition of a sale is just used to define the state; but demand depends on actual prices.

In the final set of columns we continue to define a sale as any price below one dollar but assume rational expectations instead of perfect foresight. As we discussed in Section IIA5, one way to estimate the rational expectations model is by restricting the sample to periods in which predicted demand does not depend on future prices. Since the perfect foresight and rational expectations models deliver the same predictions in the states where future price expectations do not matter, the only difference relative to columns 5 and 6 is the periods used. Restricting the sample to such states reduces the number of observations from 45,434 to 30,725.

Overall, the results from all three models are similar. For the purpose of computing the benefits from price discrimination the key is the heterogeneity in the price sensitivity. The three models suggest almost identical numbers: non-storers are significantly less price sensitive than storers. This is consistent with price discrimination being a motivation for the existence of sales. The main difference across the

\textsuperscript{11} In addition to the example in the text, the other four states where the models differ in their predictions for Coke purchases are when the Coke state is $NS$ and the Pepsi state is either $NN$ or $SN$, or the Coke state is $SS$ and the Pepsi state is either $NN$ or $SN$. Symmetric arguments hold for Pepsi.

\textsuperscript{12} Predictions differ, for example, when $p_t = 0.99$ and $p_{t+1} = 0.95$. In the second model, consumers purchase for future consumption at $t$ while in the former they wait for the better price at $t + 1$.

\textsuperscript{13} The main reason to present the results in columns 5 and 6 is to separately show the effect of the change in the definition of a sale and the effect of the change in the sample, which we show in columns 7 and 8 when we assume rational expectations.
three sets of results is in the cross-price elasticity of storers. The lower cross-price effects under perfect foresight may suggest that Assumption 4 introduces measurement error. Recall that the key complication of the rational expectations assumption was future cross-price effects. Eliminating periods where the assumption perhaps fails raises the cross price estimates. For the calculations below, the difference between the models is not of great importance.

The parameter $\omega$ measures the relative intercepts of the demand function of the two consumer types. This is not a measure of the relative importance of the two groups. Since storers are more price sensitive they will comprise a smaller fraction of demand at actual prices. Indeed, as we will see below for most observed prices, demand from non-storers will constitute the majority of quantity sold.

The own-price elasticity implied by the estimates from the dynamic model, evaluated at the quantity-weighted price, is 2.16 for Coke and 2.78 for Pepsi. The elasticities implied by the static estimates are 2.46 and 2.94, respectively. As expected, neglecting dynamics in the estimation overstates own-price elasticities.

2. Sensitivity Analysis: Heterogeneous Storage: Until now we assumed $T = 1$. We now examine the sensitivity of the results to the definition of $T$. The first two sets of columns in Table 4 present estimates of the same model as in columns 3 and 4 of Table 3, but assuming $T = 2$ and $T = 3$, respectively. The results are very similar to those in Table 3. The main change is that the estimates compensate somewhat for the storers’ extra periods of storage by slightly increasing the fraction of non-storers and the price sensitivity of storers.

Since the results are similar we want a way to choose between the different values of $T$. The last two sets of columns do this. We allow for three types: non-storers,
consumers who can store for \( T = 1 \) (denote their fraction \( \omega_{T=1} \)), and consumers who can store for longer, either \( T = 2 \) or \( T = 3 \). We look at the fraction of each type of consumers.

In columns 5 and 6 the fraction of non-storers is 0.15, essentially identical to the estimates in Table 3. The fraction of consumers who can store for \( T = 1 \) periods is 0.84, suggesting that roughly 1 percent of consumers store for \( T = 2 \). Thus, the \( T = 2 \) type are non-significant relative to the \( T = 1 \) type. The last pair of columns paints a similar picture. Here the fraction of consumers who store for three periods is less than 5 percent. These results suggest that \( T = 1 \) is the preferred option.

**B. Implications for Pricing and Welfare**

We now examine the implications of the estimates. We focus on the role of sales as a form of intertemporal price discrimination and their welfare implications. We consider two benchmarks: (infeasible) third-degree discrimination, targeting storers and non-storers with different prices, and nondiscrimination, a single price for all consumers.

This analysis neglects the vertical relationship between manufacturers and retailers, which could generate double marginalization. The relationship between retailers and manufacturers is quite interesting and subtle, but beyond the scope of this paper. The first-order conditions in equation (6) represent either a manufacturer selling to a competitive retailing industry or integrated pricing with transfers (which avoids double marginalization).

\[ \text{PF-T = 2} \]

<table>
<thead>
<tr>
<th>( P_{\text{own}} ) non-storers</th>
<th>Coke</th>
<th>Pepsi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{cross}} ) non-storers</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>( P_{\text{own}} ) storers</td>
<td>-4.25</td>
<td>-5.47</td>
</tr>
<tr>
<td>( P_{\text{cross}} ) storers</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.31</td>
<td>0.41</td>
</tr>
<tr>
<td>( \omega_{T=1} )</td>
<td>0.84</td>
<td>0.78</td>
</tr>
</tbody>
</table>

\[ \text{PF-T = 3} \]

<table>
<thead>
<tr>
<th>( P_{\text{own}} ) non-storers</th>
<th>Coke</th>
<th>Pepsi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{cross}} ) non-storers</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>( P_{\text{own}} ) storers</td>
<td>-4.34</td>
<td>-5.29</td>
</tr>
<tr>
<td>( P_{\text{cross}} ) storers</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.31</td>
<td>0.41</td>
</tr>
<tr>
<td>( \omega_{T=1} )</td>
<td>0.84</td>
<td>0.78</td>
</tr>
</tbody>
</table>

\[ \text{PF-T = 1.2} \]

<table>
<thead>
<tr>
<th>( P_{\text{own}} ) non-storers</th>
<th>Coke</th>
<th>Pepsi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{cross}} ) non-storers</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>( P_{\text{own}} ) storers</td>
<td>-4.34</td>
<td>-5.29</td>
</tr>
<tr>
<td>( P_{\text{cross}} ) storers</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.31</td>
<td>0.41</td>
</tr>
<tr>
<td>( \omega_{T=1} )</td>
<td>0.84</td>
<td>0.78</td>
</tr>
</tbody>
</table>

\[ \text{PF-T = 1.3} \]

<table>
<thead>
<tr>
<th>( P_{\text{own}} ) non-storers</th>
<th>Coke</th>
<th>Pepsi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{cross}} ) non-storers</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>( P_{\text{own}} ) storers</td>
<td>-4.34</td>
<td>-5.29</td>
</tr>
<tr>
<td>( P_{\text{cross}} ) storers</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.31</td>
<td>0.41</td>
</tr>
<tr>
<td>( \omega_{T=1} )</td>
<td>0.84</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Notes: All estimates are from least-squares regressions. The dependent variable is the (log of) quantity of Coke or Pepsi sold at a store in a week. All columns include store fixed effects. In all columns we estimate the model with perfect foresight, the differences are in the value of \( T \) and the number of different types of consumers. Standard errors are reported in parentheses.
1. Markups and Profits: A standard exercise is to use demand estimates and a first-order condition from static profit maximization to infer markups and marginal costs. Following this approach, and using the first two columns of Table 3, we compute implied markups of 43 cents for Coke and 34 cents for Pepsi. Subtracted from a quantity-weighted average transaction price of 1.07 and 1.01, respectively, these estimates lead to marginal costs of 66 and 67 cents, respectively.

Repeating this calculation using the dynamic demand estimates reported in Table 3, but still relying on a static first-order condition, we find that the implied margin for Coke is 50 cents, with marginal costs of 57 cents. We find estimates of 37 and 64 cents, respectively, for Pepsi. The lower dynamic elasticities translate into higher implied markups.

Naturally, demand dynamics render the static first-order conditions inadequate as a description of seller behavior. We therefore turn to the dynamic pricing model. We compute prices and profits under non-discrimination, under third-degree discrimination and under sales. We assume that the marginal cost during sale and non-sale periods is the same. Since each first-order condition delivers a different marginal cost, we use the average across the regimes to compute prices and profits.

Table 5 displays the optimal prices under the different pricing assumptions as well as profits relative to discrimination. By discrimination we mean the case where the firms can identify storers and non-storers, set different prices for each group, and prevent arbitrage. This is of course infeasible but serves as a benchmark to measure the maximum attainable gains from price discrimination.

By comparing the discriminatory and non-discriminatory prices we see the potential role of sales in targeting price sensitive buyers with a lower price. The optimal non-discriminatory price is 1.11 for Coke and 1.04 for Pepsi. These prices are, naturally, between the discriminatory prices for both products: 1.31 and 0.83 for Coke and 1.14 and 0.86 for Pepsi. Non-storers, being less price sensitive, are targeted with higher prices than storers. The discriminating prices target non-storers with 58 percent higher Coke prices than storers’ prices. The gap is 33 percent for Pepsi.

Rows labeled 3 and 4 present prices and profits under two different models of sales. The numbers in row 3 use the demand estimates from columns 3 and 4 of Table 3 (\( T = 1 \) and perfect foresight). In row 4 we present, for robustness, the results for the \( T = 2 \) model with perfect foresight (columns 1 and 2 in Table 4). In all cases we assume that the competitor charges the non-discriminatory price. The exercise amounts to evaluating the impact of different pricing strategies, taking as given the competitor’s behavior. In the next section we evaluate equilibrium regimes where both players discriminate.

The optimal sale price is between the non-discriminatory price and the storers’ discriminating price. It differs from the non-discriminatory price because it targets a population with a higher proportion of storers, since they purchase for two periods. By placing more weight on the price sensitive buyers, the sale price is lower than the non-discriminating one. The estimates imply a sale price about 8 percent below the non-discriminatory price for Coke, and 5 percent lower for Pepsi.

The column labeled profit displays the fraction of the discriminating profits (the highest the seller can get) accrued without discrimination and through sales, respectively. For example, for Coke the non-discriminatory seller gets 88 percent of the
discriminating profits, while sales allow the seller to accrue about 91 percent of the discrimination benchmark.

In row 4 we examine the impact of a longer storage horizon, $T = 2$. In this model sales are deeper, as they are aimed at an aggregate demand that places more weight on storers who purchase for current consumption as well as for the coming two periods ahead (as opposed to just one period ahead as in the $T = 1$ model). In turn, sales more effectively capture additional gains from discrimination, 95 percent and 97 percent of the target for Coke and Pepsi, respectively. Thus, sales allow sellers to capture around 50 percent of the gap in profits between discrimination and non-discrimination.

2. Welfare: The welfare consequences of third-degree price discrimination have been studied since Robinson (1933). The standard intuition is that price discrimination by a monopolist yields lower prices in the weak market, where the demand is more price-sensitive, and higher prices in the strong market, relative to the non-discriminatory price. So while the seller is better off, some consumers are better off and others worse off. The overall impact of discrimination is an open question subsequently studied by Schmalensee (1981); Varian (1985); and Aguirre, Cowan, and Vickers (2010) among others. A necessary condition for welfare to improve is that quantity sold increases. Since the allocation of goods across markets is distorted, a constant (or lower) output would necessarily lead to lower total surplus (for a formal proof see Schmalensee 1981).

In the context of a duopoly the picture is slightly more complex, as it is not even clear that sellers are better off under discrimination. Few papers provide theoretical results. Borenstein (1985) and Holmes (1989) offer conditions for output to increase under duopoly, and simulations showing that profits may decline. Corts (1998) shows that even with well-behaved profit functions all prices can decrease when the weak market of one firm is the strong market of the other.

---

Table 5—Gain from Sales

<table>
<thead>
<tr>
<th>Pricing</th>
<th>Coke</th>
<th>Pepsi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regular ($)</td>
<td>Sale ($)</td>
</tr>
<tr>
<td>1. Non-discrimination</td>
<td>1.11</td>
<td>88</td>
</tr>
<tr>
<td>2. Discrimination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-storers</td>
<td>1.31</td>
<td>100</td>
</tr>
<tr>
<td>Storers</td>
<td>0.83</td>
<td>100</td>
</tr>
<tr>
<td>Sales $T = 1$</td>
<td>1.31</td>
<td>1.02</td>
</tr>
<tr>
<td>Sales $T = 2$</td>
<td>1.27</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Notes: Computed based on the estimates of columns 3 and 4 of Table 3, and columns 1 and 2 of Table 4 (for the $T = 2$ case). The columns labeled Regular and Sale present the regular and sale price, respectively. The column labeled Profit is the percent profit in each regime relative to profits under discrimination. The marginal cost used in each case is computed using first order conditions averaged across different states. The imputed marginal costs are 0.60 for Coke and 0.67 for Pepsi.

---

The direction of price changes indeed follows this pattern if the monopolist’s profit function is strictly concave in price within each segment. When this is not the case the direction of price changes is ambiguous (Nahata, Ostaszewski, and Sahoo 1990).
We now evaluate the impact of intertemporal discrimination on quantity and welfare. We first consider the implication of our estimates for a monopolist (equivalently, a duopolist that unilaterally best responds given competitor behavior), and compare the findings to the theoretical literature. We then consider the duopoly case where there is little theoretical guidance.

**Best Responses.**—As a first step we compute quantity changes holding fixed the behavior of competitors and assuming that these competitors do not price discriminate. This allows us to isolate the impact of different pricing strategies. This exercise is akin to Schmalensee (1981) and Aguirre, Cowan, and Vickers (2010) because the seller is basically a monopolist. Their results apply to the intertemporal price discrimination by reinterpreting demand during sale periods, \( Q_{NS}(p) + 2Q_{S}(p) \), as the weak market, and demand during non-sale periods, \( Q_{NS}(p) \), as the strong market.

Before looking at the numbers in Table 6 we turn to the theoretical literature for predictions and to make sure that the functional forms we use are not responsible for our findings. Proposition 3 in Aguirre, Cowan, and Vickers (2010) encompasses our demand framework. They show that welfare depends on the relative concavity of the demand functions in the two markets. Our estimates deliver a more convex demand in the weak market, which is one of the conditions singled out in Robinson (1933) for quantity to increase under discrimination. In addition, what Aguirre, Cowan, and Vickers (2010) call the IRC condition \(^{16}\) holds for exponential demands and therefore Proposition 3 therein applies. Proposition 3 is a comparative static with respect to the degree of price discrimination.\(^{17}\) Proposition 3 shows that for our estimated demand welfare increases in the extent of discrimination and then declines: a little discrimination is welfare improving; full discrimination could deliver higher or lower welfare. In sum, even in the monopoly case the impact of discrimination

---

16 The condition requires that the ratio of the derivative of welfare with respect to price to the second derivative of profits be monotonic in price in each market.

17 They follow the analysis in Schmalensee and Holmes whereby the implications of discrimination are assessed by studying the behavior of a seller who is constrained to set prices in the weak and strong market no more then \( r \) units apart. As \( r \) increases the optimum approaches full discrimination.

---

Table 6—Quantity Effects (no PD by competitors)

<table>
<thead>
<tr>
<th></th>
<th>Coke</th>
<th>Pepsi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price</td>
<td>Quantity</td>
</tr>
<tr>
<td>Non-discrimination</td>
<td>318.00</td>
<td>161.86</td>
</tr>
<tr>
<td>Non-storers</td>
<td>1.11</td>
<td>258.42</td>
</tr>
<tr>
<td>Storers</td>
<td>1.11</td>
<td>59.58</td>
</tr>
<tr>
<td>Third-degree discrimination</td>
<td>397.44</td>
<td>184.58</td>
</tr>
<tr>
<td>Non-storers</td>
<td>1.31</td>
<td>194.92</td>
</tr>
<tr>
<td>Storers</td>
<td>0.83</td>
<td>202.52</td>
</tr>
<tr>
<td>Intertemporal discrimination</td>
<td>353.12</td>
<td>167.78</td>
</tr>
<tr>
<td>Non-storers–non-sale</td>
<td>1.31</td>
<td>194.92</td>
</tr>
<tr>
<td>Non-storers–sale</td>
<td>0.99</td>
<td>307.36</td>
</tr>
<tr>
<td>Storers–non-sale</td>
<td>1.31</td>
<td>0</td>
</tr>
<tr>
<td>Storers–sale</td>
<td>0.99</td>
<td>101.98</td>
</tr>
</tbody>
</table>

Notes: Computed based on the estimates of columns 3 and 4 of Table 3. Each entry shows the price, in dollars, or quantity, in units per week per store, or profits, in dollars per week per store, from each group under each regime.
on welfare is indeterminate. It is an empirical matter that we will evaluate with the estimates.

Table 6 shows prices, quantities, and profits under different pricing regimes for the different segments of the market. Quantities and profits are per week. Overall the table paints a clear picture. Both quantities and profits are higher under discrimination than non-discrimination. Intertemporal discrimination is in between. While intertemporal discrimination recovers about a quarter of the potential profit difference between discrimination and no discrimination, it delivers about half of the quantity increase.

It is interesting to see the breakdown by consumer segments. It seems like sales are a fairly efficient way of recovering the potential profits from the consumers who store. Sales seem to recover over 50 percent of the potential profits from this group for Coke. The overall gains in profits are smaller because profits from the non-storing group slightly decline. The same is true for the increase in quantity: it mostly comes from the storers. The non-storers end up paying almost an identical, slightly higher, quantity-weighted price relative to non-discrimination.

Equilibrium.—We now evaluate profits and consumer surplus when all competitors adhere to each regime. In other words, instead of best responses, as we evaluated in the previous section, we compare a regime that allows for discrimination to a regime where discrimination is not allowed. The idea is to capture market performance under different rules (e.g., if discrimination was not allowed or infeasible).

The first step is to check for quantity increases. Absent these, welfare is bound to decline. As Tables 7 and 8 show, for both products quantity increases under either form of discrimination (third-degree and intertemporal). Notice that relative to Table 6, which evaluated unilateral discrimination—competitors pricing was taken as given—quantity changes are more modest. All increases are lower, and the decline in the strong market is smaller as well. Quantity effects are attenuated by the prices of the competitor, who also discriminates.

The impact on profits is similar. Aside from the strong market in which interaction increases profits, overall profit gains are attenuated by the competitor also discriminating.

As expected, buyers in the strong market are worse off, while those in the weak market are better off. The column labeled ΔCS displays the change in consumer surplus, measured by the equivalence variation, relative to the non-discrimination case. In the case of third-degree discrimination, non-storers are worse off but storers are better off because they are offered lower prices. In the case of intertemporal discrimination, non-storers benefit during sale periods but are charged higher prices during non-sale periods; overall they are worse off. Storers are better off in both cases, but less so under intertemporal discrimination because the prices they are charged are not as low as under third-degree discrimination.

18 The equivalence variation (in this case identical to the compensating variation due to quasilinearity) of a change in two prices is the sum of the area under each demand curve as the respective prices change. That is, the area under the Coke demand curve fixing the initial Pepsi price, plus the area under the Pepsi curve fixing the final Coke price.
Total consumer welfare is lower under third-degree discrimination: the gains to the storers are outweighed by the losses of the non-storers. In the case of sales, the results differ between the products. For Coke consumer welfare slightly increases, while for Pepsi it slightly decreases. In both cases the non-storers are worse off by roughly the same amount. The difference is in the gains to storers: they are larger in the case of Coke because sales are deeper relative to non-discrimination.

Total profits increase under both forms of discrimination. Under third-degree discrimination profits from both segments increase, while under sales profits from non-storers decrease.

Total surplus is higher under discrimination of either sort than under a single price. Looking at total surplus by segment, for both products it decreased in the non-storers segment.
V. Concluding Comments

We study the impact of price discrimination when consumers can store for future consumption. To make the problem tractable we offer a simple model to account for demand dynamics due to consumer inventory behavior. We estimate the model using store-level scanner data and find that consumers who store are more price sensitive. This suggests that intertemporal price discrimination can potentially increase profits, which we then quantify. We find that sales allow sellers to recover 24–30 percent of the potential gains from (non-feasible) third-degree price discrimination. The estimates also suggest that total welfare increases when sales are offered.

A key to making our model tractable is the simplicity of the demand model. It is important to note that in order to take a more general inventory model to the data, one needs to also make some strong (mostly untestable) assumptions regarding, for example, the functional form of inventory cost and price expectations (see, for example, Erdem, Imai, and Keane 2003 or Hendel and Nevo 2006b). Whether it is more reasonable to make the assumptions of this paper or those of the previous literature depends on the application. If one is willing to make the assumptions herein, the analysis is significantly simpler, making the supply side tractable.

Appendix: Technical Details

In this Appendix we provide some technical details discussed in various sections in the text.

A. Rational Expectations

In Section IIA5 we propose a way to solve the model under rational expectations for a subset of the states. We now discuss a computationally more intensive way to solve the model in all states.

As we discussed in the text, when more than one product is storable the following issue arises. At the time of purchasing product $j$ the (effective) prices of other products may not yet be known. They would be known, for example, if the other products are on sale, in which case the effective price is the current price. In general, when some other product is not currently on sale, the consumer has to purchase product $j$ under uncertain future $-j$ prices. Thus, in period $t$ storers chose $q_{jt+1}$ to maximize expected utility

$$E_t\{u^S_{t+1}(q_{jt+1}, q^*_{jt+1}(p_{jt+1}^x, q_{jt+1}^x)) - q_{jt+1}p_{jt} - q^*_{jt+1}(p_{jt+1}^x, q_{jt+1}^x)' \ p_{jt+1}^x\},$$

where the expectation is taken with respect to the distribution of $-j$ prices. $q^*_{jt+1}(p_{jt+1}^x, q_{jt+1}^x)$ represents optimal $-j$ consumption in $t+1$ given price realizations and pre-stored quantity, $q_{jt+1}$. At $t$ the consumer chooses product $j$ purchases knowing $-j$ consumption is contingent on eventual $-j$ prices.

With a general price support the problem is tedious since it requires solving for $q^S_{jt}$ for every price in the support. However, if we assume a two-price support, $p_{jt}^N$ and $p_{jt}^S$, the problem is tractable.
B. Proof of Proposition 1

PROOF:
First consider the last two periods in isolation. As we show above, under condition (7) optimal prices are $p$ followed by $\bar{p}$.

Now consider a four-period problem. The proposed cycle is feasible and attains twice the profits of the two-period problem. To see that it is feasible notice that the candidate prices do not generate storage between periods 2 and 3. Absent a link between periods 2 and 3, consumers behave as prescribed in each independent two-period cycle (the first cycle is periods 1 and 2, the second is periods 3 and 4).

It remains to be shown that no other price sequence is more profitable. First notice that among all price sequences with $p_2 > p_3$ only our candidate $[p, \bar{p}, p, \bar{p}]$ can be optimal. Since $p_2 > p_3$ implies no storage in period 2, thus the two cycles are detached. The solution to the detached problems involves $p$ followed by $\bar{p}$ in each part.

It remains to rule out price sequences that involve storing in period 2; that is, sequences with $p_2 \leq p_3$. If $p_2 \leq p_3$ consumers store in period 2. Having purchased in period 2 for consumption in period 3, in period 3 storers would only purchase for period 4 consumption, and they would do so only if $p_3 \leq p_4$. Suppose that indeed storers purchase in period 3 for period 4 consumption. Since storers are absent in period 4, the optimal $p_4$ is $\bar{p}$ while optimal third period price is $p^*_{ND}$. In the alternative case, in which storers do not purchase in period 3, the optimal price in period 3 is $\bar{p}$ followed by $p^*_{ND}$ in period 4, as in the last period all customers are present (and there are no further purchases for storage). We now return back to the first two periods. There are two cases to consider, either storers store in period one or they do not. If they store the only candidate for first period price is $p$. Moreover, in period 2 they store for period 3. Thus the optimal price is $p^*_{ND}$. If they do not store in period 1 the optimal first period price is $p^*_{ND}$, while the second period optimal price is $p$.

Thus, the optimal prices in the first two periods (given storage between periods 2 and 3) are: $\{p^*_{ND}, p\}$ or $\{p, p^*_{ND}\}$ followed by $\{\bar{p}, p^*_{ND}\}$ or $\{p^*_{ND}, \bar{p} \}$ in periods 3 and 4. Either way profits amount to $2\pi_{ND} + \pi_{IPD}(p)$ which by condition (7) is lower than $2\pi_{IPD}(p)$. Thus, the cycling prices do better than any other price sequence in the four period problem.

We can keep adding two periods at a time and apply the same reasoning. The cycle is feasible and optimal assuming no storage between cycles. On the other hand, a sequence that induces storage between cycles, as above, fails to exploit discrimination at least once.

Finally, we need to consider odd $R$. It is easy to see that as we go from two to three periods the best the seller can get is $\pi_{ND} + \pi_{IPD}(p)$; namely, one event of discrimination plus a non-discriminating profit flow. The same is true for any $R$-long horizon.

C. Estimating Equations

We choose the parameters to minimize the sum of squares of the difference between observed purchases and those predicted by the model. Let $x_{jst}$ denote the purchases of product $j$ in store $s$ at week $t$. By (4), modifying the indicator functions
in (4) to ignore ties in effective prices (for simplicity of presentation), the purchases predicted by the model are given by

\[
x_{jst} = q_{jst}^{NS} + x_{jst}^{S} = Q_{jst}^{NS}(p_{jst}, p_{jst-s,t}) + \sum_{\tau=0}^{T} Q_{jst+\tau}^{S}(p_{jst}, p_{jst-s,t+\tau}) \mathbb{1}[p_{jst} = p_{jst-s,t+\tau}].
\]

In the case of \(T = 1\) the predicted purchases consist of three components: the purchases by non-storers and the purchases by storers for consumption at \(t\) and at \(t + 1\). Depending on the state, one or both of the components of demand by non-storers can be zero (see equation (5)). We want to stress that in all cases we use the actual prices. The definition of sales is only used to classify the states: prices are never changed.

The data consist of a panel of quantities and prices in different stores. Different stores operate at different scales and therefore attract a different number of customers and sell different average amounts. We account for this with a store fixed effect, which varies by brand. We need to transform the predicted purchases because purchases are scaled differently in different states in order to account for store fixed effects. Given the functional form in equation (9) and assuming the fraction of non-storers is given by \(\omega\),

\[
Q_{jst}^{NS}(p_{jst}, p_{jst-s,t}) = \omega e^{\alpha_{jst}} e^{-\beta_{jst}^{NS} p_{jst} + \gamma_{jst}^{NS} p_{jst-s,t}} e^{\epsilon_{jst}}
\]

\[
Q_{jst+\tau}^{S}(p_{jst}, p_{jst-s,t+\tau}) = (1 - \omega) e^{\alpha_{jst}} e^{-\beta_{jst}^{S} p_{jst} + \gamma_{jst}^{S} p_{jst-s,t+\tau} + \epsilon_{jst+\tau}}.
\]

Denote by

\[
Q_{jst}^{*NS} = \omega e^{\alpha_{jst}} e^{-\beta_{jst}^{NS} p_{jst} + \gamma_{jst}^{NS} p_{jst-s,t}} e^{\epsilon_{jst}} \quad \text{and} \quad Q_{jst+\tau}^{*S} = (1 - \omega) e^{\alpha_{jst}} e^{-\beta_{jst}^{S} p_{jst} + \gamma_{jst}^{S} p_{jst-s,t+\tau} + \epsilon_{jst+\tau}}.
\]

Then

\[
x_{jst} = e^{\alpha_{jst}} \left( Q_{jst}^{*NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{*S} \right)
\]

and

\[
\log(x_{jst}) - \overline{\log(x_{jst})} = -\log \left( Q_{jst}^{NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{S}(p_{jst}, p_{jst-s,t+\tau}) \right) - \log \left( Q_{jst}^{*NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{*S}(p_{jst}, p_{jst-s,t+\tau}) \right),
\]

where \(\overline{\log(x_{jst})}\) denotes the average over weeks within a store and product. This transformation is equivalent to the standard “within” transformation, except that

\[
\overline{\log(x_{jst})} = \log \left( Q_{jst}^{NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{*S}(p_{jst}, p_{jst-s,t+\tau}) \right)
\]

depends on the parameters of the model and cannot be done prior to estimation. In other words, we cannot transform the model prior to estimation. Rather the estimation routine needs to compute the average for each value of the parameters.
REFERENCES


Borenstein, Severin.


Blattberg, Robert C., and Scott A. Neslin.

Berry, Steven, James Levinsohn, and Ariel Pakes.

Aguirre, Inaki, Simon Cowan, and John Vickers.

Borenstein, Severin.

Robinson, Joan.


