Intertemporal Price Discrimination in Storable Goods Markets

September 8, 2012

Abstract

We study intertemporal price discrimination when consumers can store for future consumption needs. To make the problem tractable 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 anticipate purchases for future consumption, and less price-sensitive consumers. We empirically quantify the impact of intertemporal price discrimination on profits and welfare. We find that sales: (1) capture 25-30% of the profit gap between non-discriminatory and (unattainable) third degree price discrimination profits, (2) increase total welfare, and (3) impact consumer welfare modestly.

1 Introduction

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

The goal of this paper is to empirically study the role of intertemporal price discrimination in storable goods markets. Temporary price reductions (sales) can be used to separate consumers based on their ability to store. We estimate preferences, and use the estimates to test whether sales can be driven by price discrimination. To evaluate the effectiveness of sales as a price discriminating tool, we compare profits using 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, the purchase decision is dynamic. Consumers weigh the benefit of buying at a lower price during the sale with the cost of storing for future consumption. To ensure the problem is tractable we develop a simple dynamic demand model, which helps in two ways. First, demand is easy to estimate using market level data. Characterizing consumer behavior does not require solving a value function: the problem is dynamic, but easy to solve. Second, the model provides a clear delineation of what sellers must observe to solve the pricing problem.

The key to the simplicity of the demand model is the storage technology: consumers can store for a pre-specified maximum number of periods at no cost. To see how this assumption simplifies the problem consider the simplest situation with a single variety of the product, consumers have perfect foresight, and one period of storage. Period t consumption can be purchased in period $t-1$ or in period t. Absent discounting and storage costs, the consumer will purchase in the period with the lowest price. The minimum of the periods t and $t-1$ prices, what we call *effective price,* is the opportunity cost of period t consumption. The consumer's 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. The usefulness of effective prices extends to more general settings.

We assume heterogeneity in storage costs. For some consumers storage cost is too high relative to the potential benefits, and therefore do not store. These consumers' purchasing behavior is static. Storers instead behave according to our dynamic model. Their purchases are a function of the effective prices, which may not coincide with contemporaneous prices.

The parameters of the model are identified from aggregate (store-level) quantities and observed prices. The intuition for identification is relatively straightforward. Consider the above example, where consumers can store for at most one period. Holding current price constant, we observe quantity purchased at t if there was a sale in $t-1$ and if there was none. 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 where prices are lower than in the following period. With additional assumptions on the error structure, in a static model, these two quantities should be the same on average. If they are not, this tells us that dynamics are at play and also helps us pin down the shape of the dynamics, like the storage length.

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, targeting storers with lower prices. To evaluate profits and consumer surplus under different pricing regimes we use the estimated demand, and respective first order conditions. Two benchmarks are considered: non-discriminatory pricing and third degree price discrimination under the assumption of no arbitrage. Profits under third degree discrimination are a non-feasible upper bound on gains from price discrimination. The benchmark is not feasible because in practice sellers cannot target storers and non-storers with different prices. We find that (non-feasible) third degree price discrimination would increase profits by 9-14% relative to non-discriminatory prices. Sales, as a form of partial discrimination, enable sellers to capture around 24-30% of the 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 (1984) and Aguirre et al. (2010), among others. The impact of discrimination on welfare is ambiguous. In oligopoly situations there are virtually no (theoretical) welfare results. We find that total welfare increases. 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.

Besides quantifying the impact of price discrimination there are other reasons to be interested in our demand model and supply side analysis. There is a long tradition in Industrial Organization of using demand estimates in conjunction with static first order conditions to infer market power. Demand dynamics render static first order conditions irrelevant. A supply framework consistent with demand dynamics is needed to infer market power.

The demand estimates suggest that dynamics are important. Neglecting these dynamics can lead to biases (Hendel and Nevo (2006b)). Own price responses are expected to be upward biased. The estimates reflect a weighted average of long run price responsiveness (dictated by the underlying preferences) and short run inventory (intertemporal) considerations. Cross price biases have been trickier to sign. The model suggests neglecting dynamics causes a downward bias in cross price effects. The reason is that standard static estimation controls for the wrong price of the competing goods. For antitrust applications, for example, assessing 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.

1.1 Related Literature

Numerous papers in Economics and Marketing document demand dynamics, specifically, demand accumulation (see Blattberg and Neslin (1990) for a survey of the Marketing literature). Boizot et al. (2001) and Pesendorfer (2002) show that demand increases in the duration from previous sales. Hendel and Nevo (2006a) document demand accumulation and demand anticipation effects, namely, duration from previous purchase is shorter during sales, while duration to following purchase is longer for sale periods. Erdem, Imai and Keane (2003), and Hendel and Nevo (2006b) estimate structural models of consumer inventory behavior. Our demand model is motivated by this literature but offers substantial computational savings and a tractable supply analysis.

Several explanations have been proposed in the literature to why sellers offer temporary discounts. Varian (1980) and Salop and Stiglitz (1982) propose search based explanations which deliver mixed strategy equilibria, interpreted as sales. Lal (1990) presents a model in which sales arise in a pure strategy Nash equilibrium of a repeated game where premium brands (tacitly) collude to fend off private label competitors (brand with non-loyal customers). Sobel (1984), Conlisk, Gerstner and Sobel (1984), and Pesendorfer (2002) present models of intertemporal price discrimination in the context of a durable good (more recently used by Chevalier and Kashyap (2011)), while Narasimhan and Jeuland (1985) and Hong, McAfee and Nayyar (2002) do so for storable goods.

Our estimates show that sellers have an incentive to intertemporally price discriminate, suggesting that sales are probably driven by discrimination motives. Incentives for sales are similar to Narasimhan and Jeuland (1985) and Hong, McAfee and Nayyar (2002) but in a somewhat different context. Hong et al. model an homogenous good sold in different stores, under unit demand, and single unit storage. Since the interest in this paper is empirical we need a framework amenable for demand estimation, that allows for product differentiation and endogenous consumption and storage levels, depending on prices.

The abovementioned third-degree discrimination theoretical results (e.g., Robinson (1933), Schmalensee (1981), and Aguirre et al. (2010)) apply to the intertemporal discrimination in our model, once we reinterpret demand during sale periods –following Robinson's terminology– as the weak demand, and demand during non-sale periods as the strong demand.

Finally, our results relate to several papers in the empirical price discrimination literature. Shepard (1991) finds that, consistent with price discrimination motives, the price gap between full and self service is higher at gas stations offering both levels of service relative to the average difference between prices at stations offering only one type of service. Verboven (1996) considers whether discrimination can explain differences in automobile prices across European countries. 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 in airlines.

1.2 Motivating Facts

Figure 1 shows the price of a 2-liter bottle of Coke in a store over a year. The pattern is typical of pricing observed in scanner data: regular prices and occasional sales, with return to the regular price. With the exceptions 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 these facts will feed into our modeling below.

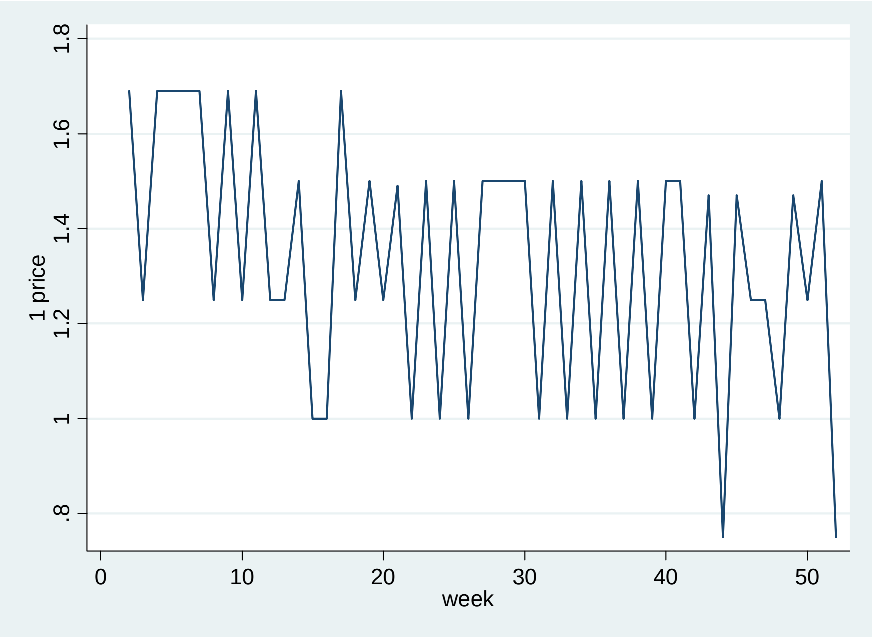

Figure 1: A typical pricing pattern

Note: The figure presents the price of a 2-litter bottle of Coke over 52 weeks in one store.

The pattern in Figure 1 raises two immediate questions: how do consumers faced with this price process behave? And what is the supply model that generates this pattern?

Since soft-drinks are storable, pricing like this creates incentives to anticipate future needs: buy during a sale for future consumption. Indeed, quantity purchased shows evidence of demand anticipation. Table 1 displays the quantity of 2-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 significantly 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).[1]

We interpret the simple patterns present in Table 1 as evidence of demand dynamics and that consumers' ability to store detaches consumption from purchases. Table 1 shows that purchases are linked to previous prices. Moreover, Table 1 hints that storing behavior might be heterogeneous. If every consumer stored during sales we would expect quantity sold at non-sale price that follows a sale to be quite small. The first row shows that is not the case.

[1] The patterns in Table 1 persist if we condition on whether or not Pepsi is on sale. Therefore, the patters we see are not driven by Pepsi price differences across states.

Table 1: Quantity of 2-Liter Bottles of Coke Sold

	$S_{t-1} = 0$	$S_{t-1} = 1$	
$S_t = 0$	247.8	199.4	227.0
$S_t = 1$	763.4	531.9	622.6
	465.0	398.9	

Note: The table presents the average across 52 weeks and 729 stores of the number of 2-litter bottles of Coke sold during each week. As motivated in the text, a sale is defined as any price below 1 dollar.

We now turn to the second question: what supply model generates such a pricing pattern? The literature has offered several explanations. First, in principle the sales pattern could be driven by changes in production costs or demand shocks. In practice, however, it's 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 in equilibrium firms pricing involves mixed strategies. While these models have some attractive features, like their ability to generate random sales, they also yield some predictions that are hard to match with the data (like a continuous price support).

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 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. This model can explain the pricing patterns but it does not explain the dynamic patterns in demand 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 in to the store.

A final set of theories focus on sales as intertemporal price discrimination (Sobel, 1984; Conlisk, Gerstner and Sobel, 1984; Pesendorfer, 2002; Narasimhan and Jeuland, 1985; Hong, McAfee and Nayyar, 2002). We focus on this class of models and test whether consumer preferences and storing ability can explain sales in this market. Finding that it can, we quantify the impact of sales.

2 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 consumers' inventory problem (as in Hendel and Nevo (2006a)) who take the price distribution as given, and the seller's profit maximization problem given demand derived from the inventory problem.

Pricing, in principle, may depend on the inventory held by each of the buyers in the population as well as their preference shocks. Given pricing, consumers purchasing decisions would depend on such information as well, should it be available to them. However, consumption and thus inventories are private information. Sellers may use the past purchases or total quantity sold, to infer inventory holdings but probably even such limited information is not available to buyers and competitors, so it is unclear what strategies depend on.

Hong, McAfee and Nayyar (2002) provide a model in which past prices, which are public information, suffice to infer past purchases. In their model preferences are deterministic and identical, 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 et al., that retains the sufficiency of public information. We want the model to allow for: 1) differentiated products, 2) consumption and the quantity stored to depend on price and 3) non-deterministic demand.

2.1 A Simple Demand Model

In this section we present a model that accommodates the empirical requirements, yet past prices are a sufficient statistic for the state. Although simple, it delivers an empirical framework for estimation. We present the main assumptions, and discuss the implied purchasing patterns.

2.1.1 The Setup

To study pricing incentives in storable goods markets we make assumptions on three components: (i) the nature of consumer heterogeneity; (ii) storage costs; and (iii) consumer expectations. First, we make the following assumption regarding consumer heterogeneity.

Assumption A1: a proportion of consumers do not store.

One can view A1 as an assumption on the distribution of storage costs, so the fraction of consumers that do not store can depend on prices. Taken literally, A1 implies that some buyers have no storage ability, are not shoppers or simply are transient buyers.

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

$$U_t^S(\mathbf{q}, m) = u_t^S(\mathbf{q}) + m \text{ and } U_t^{NS}(\mathbf{q}, m) = u_t^{NS}(\mathbf{q}) + m \quad (1)$$

where $\mathbf{q} = [q_1, q_2, ..., q_J]'$ is the vector of quantities consumed of the J varieties of the product, m is the quantity consumed of a numeraire. We index preferences by t to allow for changing needs. We assume that consumers know their future needs at least T periods in advance. 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. Heterogeneity drives sellers' incentives to discriminate. It is an empirical matter if, 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 discriminatory motives are not behind sales.

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

$$\mathbf{q}_t^S = Q_t^S(\mathbf{p}_t) \text{ and } \mathbf{q}_t^{NS} = Q_t^{NS}(\mathbf{p}_t) \quad (2)$$

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

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

A2 (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, at the store buyers only consider purchasing for T periods ahead. The assumption delivers simple dynamics; it simplifies the state and detaches storage decisions of the different varieties. As we will see below, under A2, effective prices, of the other products, are a sufficient statistic for quantity in storage. Effective prices, defined later, are public information. Thus, the assumption not only eases demand estimation but also helps formulate the sellers' problem.

Finally, we assume the following about price expectations.

A3 (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. One saving grace is that once we solve the supply model then pricing will be deterministic and easy to predict. Therefore, the model with perfect foresight is internally consistent.

We also consider an alternative assumption:

A3' (rational expectations): consumers hold rational expectations about future prices.

We present estimates under both A3 and A3' for $T = 1$, and the results are similar.[2] For larger T estimation remains trivial under perfect foresight, while under A3' it gets quite complicated. Unlike future prices, which A3' allows to be uncertain, we continue to assume throughout perfect foresight of future needs T periods ahead. For low T this seems like a reasonable assumption.

2.1.2 The Storing Consumer's Problem

Aggregate purchases are the sum of the purchases of the two types of consumers:

$$X_{jt}(\mathbf{p}_{t-T}, .., \mathbf{p}_{t+T}) = Q_{jt}^{NS}(\mathbf{p}_t) + X_{jt}^{S}(\mathbf{p}_{t-T}, .., \mathbf{p}_{t+T}) \tag{3}$$

where $X_{jt}^{S}(\mathbf{p})$ are purchases by storers. Purchases by non-storers are given by static demand, (2). Note that since the utility needs vary over time the function X is indexed by t. A private case, which we will use in the empirical analysis, will be $X_{jt}(\cdot) = X_j(\cdot) + \varepsilon_{jt}$, where ε_{jt} is a time varying shock.

We first find storers' demand under assumptions A1-A3, and later consider A3'. Under perfect foresight the consumer knows in which of the T preceding periods it is the cheapest time to purchase for period t consumption. We define the *effective prices* as the minimum price in the relevant $T+1$ 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. With many products and potential ties, deriving purchases requires some notation.

For ease of exposition we ignore discounting. The application involves weekly data, and therefore discounting does not play a big role. In order to insure a solution to the dynamic problem we focus on a finite horizon, up to period R.

[2] We believe the reason estimates are similar is that there are well defined sales and non-sale prices. If consumers do not know future prices, but there are two clear price ranges – of sale and of non-sale prices – then they purchase for storage on sale, and never store at non-sale prices. In other words, under a two-price regime knowledge of future prices is not needed to generate the same storing behavior that arises under perfect foresight. Although in the application there are more than two prices, in Figure 1 one can see a clear distinction between sale and non-sale prices.

Storers maximize the sum of utility over the R periods subject to a budget constraint and the storing constraints:

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

$$\sum_{t=1}^{R} (\mathbf{p}_t' \mathbf{x}_t + m_t) \leq \sum_{t=1}^{R} y_t \text{ and } \mathbf{q}_t \leq \mathbf{x}_t + \sum_{\tau=1}^{t-1} (\mathbf{x}_\tau - \mathbf{q}_\tau - \mathbf{e}_\tau)$$

where $\mathbf{x}_t$ is the vector of purchases, and $\mathbf{e}_\tau$ is the vector of unused units that expire in period τ. 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[3]

$$p_{jt}^{ef} = \min\{p_{jt-T}, .., p_{jt}\} \tag{4}$$

Notice the effective price is actually the opportunity cost, since items will not be purchased at any other price than the effective price. Denote by $\mathbf{p}_t^{ef}$ the vector of period t effective prices, the problem of the storing consumer becomes

$$Max \sum_{t=1}^{R} \mathbb{E}(u_t^S(\mathbf{q}_t) + m_t) \text{ subject to } (\mathbf{p_t^{ef}}\,' \mathbf{q}_t + m_t) \leq y_t.$$

The optimization of the storer is equivalent to R static optimization problems, where in each of them prices are $\mathbf{p_t^{ef}}$. Optimal consumption in period t is $\mathbf{q}_t^S = Q_t^S(\mathbf{p_t^{ef}})$. 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 taken care simply by replacing prices by 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 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. How do inventoried units of j affect demand for $-j$? Since p_{jt}^{ef} was known at the time of purchasing all other varieties to be consumed at period t, the impact is fully captured by p_{jt}^{ef}.

[3] We assume for $T > t$ $p_{jt-T} = \infty$, namely, purchases before period 0 are not feasible.

2.1.3 Purchasing Patterns

Predicted consumption is the solution to a sequence of static problems, where prices are replaced by effective prices. Predicted purchases are immediately implied, but require a bit of accounting and some unpleasant notation (mostly to take care of tie-breaks).

At any given period t storers can purchase for consumption during the current period and up to T periods ahead. Predicted purchases by storers, in period t, are the sum over current and future consumption needs supported 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 $[p_{jt-T+1}, ..., p_{jt+1}]$ 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 $\mathbf{p}^{ef}_{-j,t+\tau}$, where $\tau = 0, ..., T$. Note, that the effective price varies by -consumption- period since the other products consumed at $t+\tau$ might actually be purchased at different times.

The rest of the section formalizes this idea. To define purchases more precisely we need to take care of tie-breaks in prices, which requires additional notation. Readers are advised to move on to the next section, unless interested in tie-breaks.

In period t the consumer might purchase for t as well as for some or all of the following T periods. For $r = 0, ..., T$ purchases at time t for consumption at time $t+r$ equal either 0, if $p_{jt} > p^{ef}_{jt+r}$, or (absent ties in effective prices) $Q^S_{t+r}(\mathbf{p^{ef}_{t+r}})$, if $p_{jt} = p^{ef}_{jt+r}$.

However, since prices may repeat themselves between periods $t-T$ and t consumers are indifferent when to purchase. We break the tie by assuming that buyers purchase immediately when the price is below a threshold, p^m_j, and wait otherwise. Namely, for $p^{ef}_{jt} < p^m_j$ consumers buy right away, while for $p^{ef}_{jt} \geq p^m_j$ consumers buy in the last opportunity in which prices equal p^{ef}_{jt}. The threshold p^m_j triggers action. A possible rationale for this arbitrary tie-breaking rule is a little uncertainty about either going to the store or about future prices. As a practical matter we can use median price as the threshold.

Define $t^f = \min\{\arg\min_{\tau \leq T}\{\mathbf{p}_{jt}, .., \mathbf{p}_{jt-\tau}\}\}$ and $t^l = \max\{\arg\min_{\tau \leq T}\{\mathbf{p}_{jt}, .., \mathbf{p}_{jt-\tau}\}\}$. These are the first and last time p^{ef}_{jt} is charged in the $t-T$ to T period, respectively. Period t purchases are:

$$X^S_{jt}(\mathbf{p}_{t-T}, ..., \mathbf{p}_{t+T}) = \sum_{r=0}^{T} Q^S_{jt+r}(p_{jt}, \mathbf{p}^{ef}_{-j,t+r})\mathbf{1}\{((t = t^f) \cap (p_{jt} < p^m_j)) \cup ((t = t^l) \cap (p_{jt} \geq p^m_j))\} \tag{5}$$

where $Q^S_{jt+r}(\cdot)$ is static demand defined by (2). The indicator function takes care of two requirements. First, we only want to purchase at time t for consumption in period $t+r$ if $p_{jt} = p^{ef}_{jt+r}$, that is, if price t is as good as any other price between $t+r-T$ and $t+r$. Second, the indicator takes care of breaking ties. Purchases take place at t only if t is the first event in which a price below the threshold is offered, or the last, if the price is above the threshold.

2.1.4 Predicted Behavior for T=1

We now spell out what equation (5) implies in the case of $T=1$. Focusing on $T=1$ serves two purposes. First, it clarifies the exact working of the model, which will help in discussing identification in the next section. Second, it prepares the ground to show how the model works under rational expectations.

Equation (5) determines whether units for consumption, by storing consumers, in periods t up to $t+T$ are purchased in period t. For $T=1$ it boils down to two decisions: (i) whether (current) period t consumption is purchased at t or was it already purchased at $t-1$, and (ii) whether the consumption for period $t+1$ is purchased in period t or in period $t+1$.

For the purpose of estimation it is convenient to translate equation (5) into a slightly different form. Define sale periods, S, 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, is one which $p_{jt} > p_{j,t+1}$.[4]

Thus, to predict storers' behavior we only need to define 4 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).

Assume for a moment that only product j can be stored. Given assumptions A1-A3 storers' purchases at time t are given by

$$x^S_{jt}(\mathbf{p}_{t-1}, \mathbf{p}_t, \mathbf{p}_{t+1}) = \begin{cases} Q^S_{jt}(p_{jt}, p_{-jt}) & & 0 & \text{in } NN \\ 0 & & 0 & \text{in } SN \\ Q^S_{jt}(p_{jt}, p_{-jt}) & + & Q^S_{jt+1}(p_{jt}, p_{-jt+1}) & \text{in } NS \\ 0 & & Q^S_{jt+1}(p_{jt}, p_{-jt+1}) & \text{in } SS \end{cases} \tag{6}$$

where $Q^S_{jt}()$ is the static demand of storers (defined in (2)).

At high prices there are no incentives to store, in which case purchases equal either: consumption, given by $Q^S_{jt}()$, 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

[4] If $p_{jt} = p_{j,t+1}$ then the period is classified as a sale if p_{jt} is below a threshold p^m_j and a non sale otherwise.

sale, current consumption comes from stored units, so purchases are for future consumption only, $Q^S_{jt+1}(p_{jt}, p_{-jt+1})$. Notice the difference in the second argument of the anticipated purchases relative to purchases for current consumption (e.g., SS vs NN). Purchases for future consumption take into account the expected consumption of products $-j$.

When all products are storable accounting for storability is immediate. The way to incorporate the dynamics dictated by storage is to replace p_{-jt+1} by p^{ef}_{-jt+1}, as discussed in the previous section. For example, consider the event NN (product j is not on sale at t nor at $t-1$) and assume that product $-j$ was on sale at $t-1$ (but is not on sale at t). Storers' purchases are $Q^S_{jt}(p_{jt}, p^{ef}_{-jt})$ instead of $Q^S_{jt}(p_{jt}, p_{-jt})$. A similar adjustment is needed in every state, to account for consumption out of storage.

There are two important implications of equation (6) for estimation. First, equation (6) dictates in which states to scale demand up or down to account for storage and consumption out of storage. The definition of N and S is used to classify the state but prices are not restricted to take on two values, as equation (6) suggests we will use actual prices for estimation.

Second, we can see from equation (6) 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 the 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 (6) requires adjustment to reflect that consumers can anticipate for longer periods ahead, as reflected in (5), basically rescaling up in case of anticipated purchases, and rescaling demand down, in case of consumption out of storage. Moreover, effective prices are the lowest of the T preceding prices. Period t consumption is the static demand based on $\mathbf{p^{ef}_t}$, while purchases take place at the time the effective price is charged.

2.1.5 Rational Expectations

We now explore demand under A3', rational expectations of future prices, while continuing to assume A1, A2, $T = 1$ and that future consumption needs are known. Equation (3) 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 et al. (2002) and Perrone (2010), there is a price threshold below which the good is purchased for future consumption. Above the threshold buyers do not have an incentive to store, denote these periods as N periods. During a sale period, S, consumers have an incentive to store, even if some sales are deeper than others. Thus, rational expectations and unrestricted price support work fine in the single product case.[5]

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

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

where expectation is taken with respect to the distribution of $-j$ prices. $\mathbf{q}^*_{-jt+1}(\mathbf{p}^{ef}_{-jt+1}, q_{jt+1})$ 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 $\mathbf{q}^*_{-j}$ for every price in the support. There are two approaches to simplify the problem. First, we could assume a two price support: p^N_{-j} and p^S_{-j}. Optimal purchases are characterized by three first order conditions.[6] The second approach, which we adopt below, involves assuming an unrestricted price support but using a restricted sample 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 future prices.

In those 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 4 states of each good) predicted purchases of each product are independent of future prices. In total only 2 out of the 16 states involve behavior that differs

[5] The only minor subtlety is whether consumers that stockpiled buy additional units should the realized price be lower than the preceding price. It is possible that consumers purchase additional units given 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 (3), as written, assumes the latter.

[6] For example, suppose Coke is on sale and Pepsi is not, the consumer has to decide how much Coke to purchase for $t+1$ knowing Pepsi's price at $t+1$ may end up at one of two different levels. The demand for Coke involves the solution of three first order conditions for: q^C, Coke purchase at t for consumption at $t+1$, $\overline{q}^P$, Pepsi consumption at $t+1$ if Pepsi ends up being on sale, and $\underline{q}^P$, Pepsi consumption absent a sale.

for both products. Thus, a way to estimate the model under rational expectation, without incurring additional computation costs and without the two price assumption, is to restrict the sample to periods where demand under both expectations assumptions coincide. We elaborate below.

2.2 Seller Behavior

We want to compare pricing regimes. For that, we need predicted seller behavior. We start by considering a monopolist facing the population of storers and non-storers, as described in the previous section, with flow demands $Q_t^S(p)$ and $Q_t^{NS}(p)$, respectively. We will assume the flow demands as $Q_t^S(p) = Q^S(p) + \varepsilon_t^S$ and $Q_{jt}^{NS}(p) + \varepsilon_t^{NS}$, where ε_t^S and ε_t^{NS} are *i.i.d* shocks with zero mean. These shock reflect varying needs. We assume that firms do not observe these shocks when setting prices and therefore optimal prices do not depend on these shocks. Furthermore, since 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 marginal cost, denoted by c, is constant over time. Below we consider the implications of (stochastically) varying costs. Let p_S^* and p_{NS}^* be the prices that maximize, static, expected profits from separately selling to the populations of storers and non-storers, and p_{ND}^* the optimal price of a non-discriminating monopolist facing the whole population. p_{ND}^* maximizes $(Q^{NS}(p) + Q^S(p))(p - c)$. Denote $\pi_{ND}^* = (Q^{NS}(p_{ND}^*) + Q^S(p_{ND}^*))(p_{ND}^* - c)$.

Timing and strategies At time 0 the monopolist commits to the sequence of prices p_1 to p_R, where R is the length of the horizon.[7] As before, we assume R is finite to avoid an infinite sum due to the lack of discounting. Storers, who under A3 know the price sequence, behave as described in (6). For example, for $T = 1$ demand is a function of the prices at periods $t-1$, t and $t+1$: $X_{jt}(p_t|p_{t-1}, p_{t+1}^e)$. The monopolist maximizes:

$$\sum_{t=1}^{R} \pi(p_t|p_{t-1}, p_{t+1}) = \sum_{t=1}^{R} X_{jt}(p_t|p_{t-1}, p_{t+1}^e)(p_t - c)$$

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

2.2.1 Two-period problem

Lets start with a two period set-up. The two-period problem helps highlight the gains from intertemporal price discrimination, and the conditions for discrimination to be profitable.

[7] Commitment seems an adequate assumption for our application. Supermarkets set prices several weeks in advance.

If prices are constant or declining, nobody stores. Absent storage, optimal prices are p^*_{ND} in both periods. Instead, with increasing prices storers purchase in the first period for consumption in both periods. In the second period, since only non-storers are present, the optimal price is p^*_{NS}. Increasing prices enable price discrimination. Profits as a function of first period price $\underline{p} < p^*_{NS}$ are:

$$\pi_{IPD}(\underline{p}) = (Q^{NS}(\underline{p}) + 2Q^S(\underline{p}))(\underline{p} - c) + Q^{NS}(p^*_{NS})(p^*_{NS} - c)$$

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, during the second period. $\pi_{IPD}(\underline{p})$ stands for intertemporal price discrimination profits, with first period -sale- price $\underline{p}$.

A constant price could be optimal. It would be, for example, if $p^*_{NS} \leq p^*_S$. However, if

$$\pi_{IPD}(\underline{p}) > 2\pi^*_{ND} \tag{7}$$

for some $p^*_S < \underline{p} < p^*_{NS}$ then the pair $\{\underline{p}, p^*_{NS}\}$ does better than twice non-discrimination profits. It is an empirical matter whether constant or increasing prices are optimal. It depends –in Robinson (1933)'s terminology– on whether non-storers are the strong and storers the weak market.

We assume throughout the concavity of objective functions. Thus, when condition (7) holds $\underline{p}$ and p^*_{NS} solve the following first order conditions:

$$\begin{aligned} p^*_{NS} &= c - \frac{Q^{NS}(p^*_{NS})}{\frac{\partial Q^{NS}(p)}{\partial p}\big|_{p=p^*_{NS}}} \\ \underline{p} &= c - \frac{Q^{NS}(\underline{p}) + 2Q^S(\underline{p})}{\frac{\partial (Q^{NS}(p)+2Q^S(p))}{\partial p}\big|_{p=\underline{p}}} \end{aligned} \tag{8}$$

In sum, if it is profitable to have a sale it should be in the first period. The sale helps target storers in the first period with a low price. In the second period, with the storers off the market, the seller targets non-storers with the higher price. The cost of having the sale is selling to non-storers at a price lower than p^*_{ND}, during the sale period. The benefits are selling to the non-storers at a higher price, p^*_{NS}, in the second period, and to storing consumers at price lower than p^*_{ND} for their consumption in both periods.

2.2.2 Multiple-period problem

In longer horizons, under condition (7) and commitment, the monopolist sets cycling prices. A sale period is followed by T non-sale periods, when only non-storers purchase. Price cycles of length $T+1$ maximize the discriminating opportunities.[8]

The simplest case to consider is $T = 1$; the analysis is similar for larger T. We assume there is a single-product monopolist who can commit to future prices; we later argue that the same predictions arise absent commitment, and under duopoly with commitment.

Proposition 1 *Under condition (7) optimal pricing involves cycles of* $\underline{p}$ *followed by* $\overline{p}$*, with* $p_S^* < \underline{p} < \overline{p} = p_{NS}^*$

Proof. First consider the last two periods in isolation. Under condition 7 optimal prices are $\underline{p}$ followed by $\overline{p}$.

Now consider a 4-period problem. The proposed cycle is feasible and attains twice the profits of the two period problem. To see it is feasible notice that the candidate prices do not generate storage between periods 2 and 3. Absent a link between period 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 sequence with $p_2 > p_3$ only our candidate $[\underline{p}, \overline{p}, \underline{p}, \overline{p}]$ 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 $\underline{p}$ followed by $\overline{p}$ in each part.

It remains to rule out price sequences that involve storing in period 2, i.e., sequences with $p_2 \leq p_3$. If $p_2 \leq p_3$ consumers store in period 2. Having purchased for period three consumption in period two, in period three storers would only purchase for period four consumption, and they would do so only if $p_3 \leq p_4$. Suppose that indeed storers purchase in period three for period four consumption. Since storers are absent in period four, the optimal p_4 is $\overline{p}$ while optimal third period price is p_{ND}^*. In the alternative case, in which storers do not purchase in period three the optimal price on period three is $\overline{p}$ followed by p_{ND}^* in period four, since in the last period all customers are present (and there are no further purchases for storage). Lets go 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 $\underline{p}$, moreover, in period two they store for period three, thus the optimal price is p_{ND}^*.

[8] When the length of the horizon $R > T+1$ –if discrimination is profitable- cyclical prices are optimal. It is immediate to see that for any $R \leq T+1$, where R is the length of the horizon, the analysis of the previous section is still valid, with optimal prices being $\underline{p}$ for one period -to supply storers- followed by p_{NS}^* afterwards. As long as the horizon is no longer than the storing period the seller first clears storers out of the market and then targets non-storers.

If they do not store in period one the optimal first period price is p^*_{ND}, while the second period optimal price is $\underline{p}$. Thus, the optimal prices in the first two periods (given storage between periods two and three) are: $\{p^*_{ND}, \underline{p}\}$ or $\{\underline{p}, p^*_{ND}\}$ followed by $\{\overline{p}, p^*_{ND}\}$ or $\{p^*_{ND}, \overline{p}\}$ in periods three and four. Either way profits amount to $2\pi^*_{ND} + \pi_{IPD}(\underline{p})$ which by condition (7) is lower than $2\pi_{IPD}(\underline{p})$. 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 $\pi^*_{ND} + \pi_{IPD}(\underline{p})$, namely, one event of discrimination plus a non-discriminating profit flow. The same is true for any R-long horizon. ■

The pricing cycle leads to a flow profit (per two-period cycle) of $\pi_{IPD}(\underline{p})$. The prediction is similar to Narasimhan and Jeuland (1985), which shows that cyclical pricing (sales) help sellers intertemporally price discriminate when buyers with more intense needs have more limited storage. It is interesting to notice that the cycle is designed to maximize the number of times the seller discriminates.

It is not difficult, but tedious, to show that the solution is time consistent. It involves showing 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, after a sale, p^*_{NS} is optimal. In the third period, the second cycle starts; it is easy to show the low-high strategy is optimal. In the second period the seller could be tempted to induce storage. However, it is easy to show the continuation of the game after such price involves lower profits.

2.2.3 Duopoly

In the context of our application there is more than one manufacturer. We now argue that in equilibrium, under commitment, both sellers charge cycling prices. The game involves two players, each simultaneously committing to a sequence of prices $\{p_{j1}, .., p_{jR}\}$. As before prices are set prior to observing the time varying demand needs, they are actually set at time 0 to maximize expected profits. Storers behave as in (6), basing their period t consumption on the lowest of t and $t-1$ prices, which by A3 are known in advance. We discuss the Nash equilibrium of this game.

Consider firm j taking as given firm $-j$'s cycling price sequence. Since storers always purchase good $-j$ at the sale price, their demand is $X^S_j(p_j, \underline{p}_{-j})$, namely, demand is the same in every period regardless of the actual value of p_{-jt}. In contrast, the non-storer's demand $Q_j^{NS}(p_j, p_{-j})$ depends on the contemporaneous $-j$ price.

Define j 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 set up:

$$(Q_j^{NS}(\underline{p}_j, \underline{p}_i) + 2Q_j^S(\underline{p}_j, \underline{p}_i))(\underline{p}_j - c) + Q_j^{NS}(\overline{p}_j, \overline{p}_i)(\overline{p}_j - c) > \pi^*_{ND}(\underline{p}_i) + \pi^*_{ND}(\overline{p}_i) \qquad (9)$$

The first term accounts for profits during the sale, when storers buy for both periods and non-storers for current consumption. The second term are profits from non-storers. The right hand side are non-discriminating profits, first while product $-j$ is on sale, and the second term for the competitors non-sale price. Condition (9) 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 (9) 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.[9]

3 Data and Estimation

3.1 Data

The data set we use was collected by Nielsen and it includes store-level weekly observations of prices and quantity sold at 729 stores that belong to 8 different chains throughout the Northeast US, for the 52 weeks of 2004. We focus on 2-liter bottles of Coke, Pepsi and store brands, which have a combined market share of over 95 percent of the market.

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.[10] 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

[9] There is an additional issue, the timing of the sales. It is possible that non-coincidental cycles are more profitable than coincidental sales. The issue does not arise in a two-period set up since there is no point of having a sale in the second period. We abstract for the moment from the timing issue, that could arise in a longer horizon, a condition like 9 would determine optimal timing.

[10] 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.

chain (in a given week), but prices across chains are different. 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 do not observe the actual price charged just a quantity weighted average. In principle, we could try to impute the missing prices. Since this is orthogonal to our main point we drop these chains.

Figure 2 displays the distribution of the price of Coke in the five chains we study below. For some of the results below we need to define a sale price – a price for which consumers stockpile. The distribution seems to have a break at a price of one dollar, so we define any price below a dollar as a sale, namely, a price at which storers purchase for future consumption. This is an arbitrary definition. A more flexible definition may allow for chain specific thresholds, or perhaps moving thresholds over time. For the moment we prefer to err on the side of simplicity. Using this definition we find that approximately 30 (36) percent of the observations are defined as a sale for Coke (Pepsi).

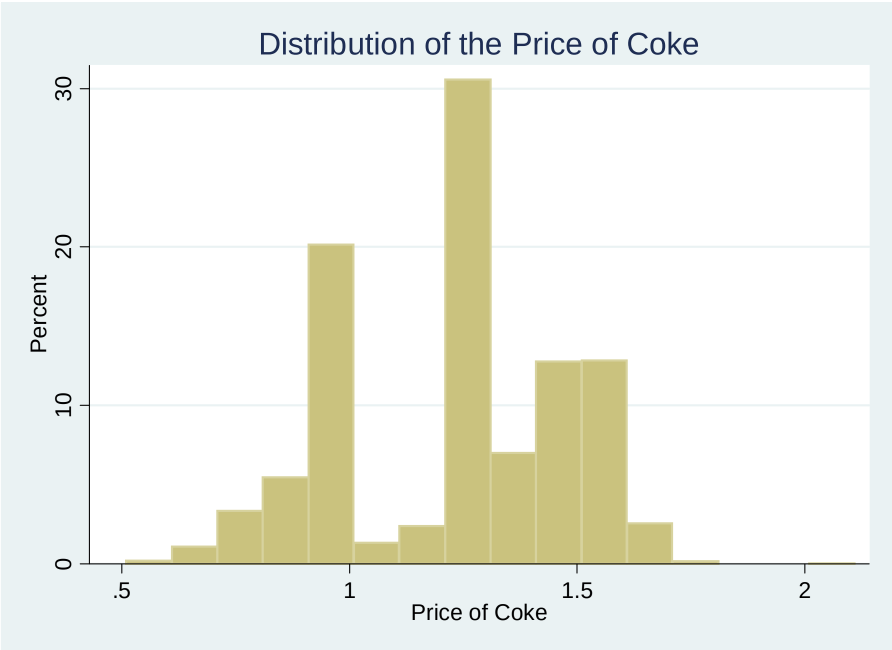

Figure 2: The Distribution of the Price of Coke

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

For the analysis below we use 24,674 observations from five chains. The descriptive statistics for the key variables are presented in Table 2.

Table 2: Descriptive Statistics

			% of variance explained by:		
Variable	Mean	Std	chain	week	chain-week
Q_{Coke}	446.2	553.2	5.6	20.4	52.5
Q_{Pepsi}	446.0	597.8	2.8	24.4	46.7
P_{Coke}	1.25	0.25	7.1	29.7	79.9
P_{Pepsi}	1.19	0.23	7.5	30.7	79.8
Coke Sale	0.30	0.46	6.4	30.0	86.6
Pepsi Sale	0.36	0.48	9.3	29.2	89.0

Note: Based on 24,674 observations for five chains, as explianed in the text. A sale is defined as any price below one dollar.

3.2 Identification and Estimation

In this section we explain how we go from the model to the estimation. The estimation is straightforward and can be preformed using off-the-shelf software (we used Stata to compute all the results below). Nevertheless, there are some issues that deserve discussion.

Overview We recover the utility parameters by matching weekly store level purchases observed in the data to those predicted by the model. The purchases predicted by the model have two components: the purchases of non-storers and those of storers. Thus, the observed purchases, X_{jt}, are

$$X_{jt} = Q_{jt}^{NS}(\mathbf{p}_t) + X_{jt}^{S}(\mathbf{p}_{t-T}, ..., \mathbf{p}_{t+T}) = Q_{j}^{NS}(\mathbf{p}_t) + X_{j}^{S}(\mathbf{p}_{t-T}, ..., \mathbf{p}_{t+T}) + \varepsilon_{jt}$$

where $Q_j^{NS}(\mathbf{p}_t)$, the purchases of non-storers, is defined by equation (2), $X_j^S(\mathbf{p}_{t-T}, ..., \mathbf{p}_{t+T})$, the purchases of storers is given by equation (5) (which in the case of $T = 1$ equals the formula given in equation (6) with effective prices replacing price for the $-j$ goods.), and ε_{jt} is an error term discussed in the next subsection. To compute the predicted purchases we need to parametrize the demand functions, Q_j^{NS} and Q_j^S. In addition, we need to define the price regimes, i.e., 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 will assume $E(\varepsilon_{jt}|\mathbf{p}_{t-T}, ..., \mathbf{p}_{t+T}) = 0$ and therefore estimate the parameters of the demand functions using non-linear 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.

First, consider the error term of the demand equation. Weekly purchases at the store level are quite volatile. In the specification below we include an additive *i.i.d.* shock, ε_{jst}, unobserved by firms at the time of pricing. ε_{jst} explains randomness in purchases given prices. The shock reflects varying needs, and is consistent with the demand model, where we assumed consumers have varying needs known T periods ahead.

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 pricing follows a deterministic cyclical pattern). While the sale and non-sale pattern is apparent in the data, it is not deterministic and typically involves more than two prices. The question is what explains the discrepancy. Note that since demand shocks are unknown at the time of pricing, they 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. On the one hand, prices determine how much is consumed. For this role, both in the theory and in the data, prices are used as a continuous variable. In addition, prices determine the timing of purchases, namely, the state. During sales storing consumers purchase for future consumption. We classify sales, or price regimes, using the history of prices. Under perfect foresight period t is defined as sale if $p_t \leq p_{t+1}$. 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. Only the price regimes, that determine which components of the storers demand switch on and off, are discrete.

For example, consider the $T = 1$ case. $X_j^S(\mathbf{p}_{t-T}, ..., \mathbf{p}_{t+T})$ is defined by equation (6). The states are defined using observed prices. Under perfect foresight, t is a sale period if p_t is lower than p_{t+1}. Once the states are defined, demand in each period is scaled up or down as in (6). The prices that enter the demand are effective prices, defined in equation (4), which are actual prices and not averages within a price regime.

Identification To see intuitively how preferences are identified, consider the case $T = 1$, for simplicity assume a single product. All the arguments go through with multiple products, but the conditioning vector needs to include effective prices of other goods. For any current price, $\mathbf{p}_t$, we observe average purchases, $X_j(\mathbf{p}_t|sale\ at\ t-1) = \sum_t X_{jt}(\mathbf{p}_t|sale$

at $t-1$) and $X_j(\mathbf{p}_t|\textit{no sale at } t-1) = \sum_t X_{jt}(\mathbf{p}_t|\textit{no sale at } t-1)$, where a "sale" is defined as explained in the previous paragraph. For example, in the case of perfect foresight, the event of "sale at $t-1$" is equal to $p_{jt-1} \leq p_{jt}$. The averaging is over time, namely, over realizations of ε_{jt}. The quantities observed for each $\mathbf{p}_t$ can be used to recover the unknown quantities $Q_j^{NS}(\mathbf{p}_t)$ and $Q_j^S(\mathbf{p}_t)$ (which in turn identify respective preferences).

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 will not purchase (they purchased at $t-1$ for consumption at t, and since p_{jt} is not a sale then they will not purchase for $t+1$). Therefore, the observed purchase equals the demand of non-storers, $X_j(\mathbf{p}_t|\textit{sale at } t-1) = Q_j^{NS}(\mathbf{p}_t)$. Furthermore, if there was no sale at $t-1$ then both storers and non- storers purchase only for current consumption therefore if p_{jt} is a not sale price then $X_j(\mathbf{p}_t|\textit{no sale at } t-1) = Q_j^{NS}(\mathbf{p}_t) + Q_j^S(\mathbf{p}_t)$. Combining with the previous result, which recovers the demand of non-storers, we get $Q_j^S(\mathbf{p}_t) = X_j(\mathbf{p}_t|\textit{no sale at } t-1) - X_j(\mathbf{p}_t|\textit{sale at } t-1)$. 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 $X_j(\mathbf{p}_t|\textit{sale at } t-1) = Q_j^{NS}(\mathbf{p}_t) + Q_j^S(\mathbf{p}_t)$ since non-storers purchase for current consumption and storers purchase for consumption at $t+1$. Similarly, $X_j(\mathbf{p}_t|\textit{no sale at } t-1) = Q_j^{NS}(\mathbf{p}_t) + 2 * Q_j^S(\mathbf{p}_t)$ since now storing consumers purchase for consumption at t and at $t+1$. This implies

$$Qj^S(\mathbf{p}_t) = X_j(\mathbf{p}_t|\textit{no sale at } t-1) - X_j(\mathbf{p}_t|\textit{sale at } t-1),$$

and

$$Q_j^{NS}(\mathbf{p}_t) = 2 * X_j(\mathbf{p}_t|\textit{sale at } t-1) - X_j(\mathbf{p}_t|\textit{no sale at } t-1).$$

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

Once we recover the demanded quantities from the observed purchases the utility parameters can be identified following standard arguments. Note, that in principle for a some sale definition and price distributions there is a price range that is sometime a sale and sometime not.[11] For these prices demand is actually over identified.

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.

[11] For example, assume that $p_{jt} = p_{j\tau}$ for periods t and τ. Furthermore, assume that $p_{jt} \leq p_{jt+1}$ and $p_{j\tau} > p_{j\tau+1}$, then using the perfect foresight definition of a sale p_{jt} is a sale price, while $p_{j\tau}$ is not.

To further understand how the data separates different storing horizons consider the following. If $T = 0$, i.e., consumers do not store, then $X_j(\mathbf{p}_t|sale\ at\ t-1) = X_j(\mathbf{p}_t|no\ sale\ at\ t-1)$ for all p_t. As we showed in the previous paragraphs, this equality does not hold under the $T = 1$ model. This gives us a way to separate and test the model. Indeed, a version of this is what Table 1 does. Intuitively, the static model says that current purchases are not a function of lagged prices, while the dynamic model allows for this 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 these longer lags matter. By conditioning on further and further lags the data can distinguish between more models.

With $T > 1$ the model is generally over identified since we have many more relevant histories to condition on. We can use this additional relevant histories to enrich the model and allow for more types. Specifically, we allow for 3 types: non-storers ($T = 0$), storers with $T = 1$, and long horizon storers with $T = 2$ (or with $T = 3$). The estimates for these models are in Section 4.1.2 (please see Table 4). We find that the fraction of longer horizon storers is small once we allow for them.

Estimation For estimation we assume demand for product j at store s in week t is log-linear:

$$\log q^h_{jst} = \omega^h \alpha_{sj} - \beta^h_j p_{jst} + \gamma^h_{ji} p_{ist} + \varepsilon_{jst}, \quad j = 1,2 \quad i = 3 - j \quad h = S,\ NS \tag{10}$$

where α_{sj} is a store specific intercept for each brand, and ε_{jt} is an i.i.d. shock. The parameters ω^h allow for different intercepts for each consumer type. We scale these parameters to add up to one and define $\omega = \omega^{NS} = 1 - \omega^S$. ω is the ratio of intercepts, which 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-storer, are similar. However, in the linear model predicted demand can be negative (specially storers' demand at a high price). The log-linear specification avoids the negativity problem by imposing an asymptote to zero consumption as price increases.

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. We present the exact estimating equations in the Appendix. We should stress that in all cases we use actual prices: the definition of sales is used to define the states but in no way do we modify or average prices within state. 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 example below. To obtain the exact estimating equations we combine equations (3) and (10), 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. Still the estimation is straightforward. 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 2.1.5, defining the purchases by storers is more complex under rational expectations. In particular, absent perfect foresight defining effective prices is problematic. We offer two solutions. The first involves assuming a two price support and solving a system of equations, defined in Section 2.1.5. The second approach, which we follow below, involves estimating the model as in the perfect foresight case, but using only part of the sample.

With 2 products and $T = 1$, there are 16 states (4 for each product). In 10 of these states the purchasing pattern predicted by the rational expectations model coincide with predicted purchases under the perfect foresight model, using the same definition of a sale, and therefore we can 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, wether 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 prescribe 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 prediction. 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, how much Coke is bought at t.[12]

4 Results

4.1 Demand Estimates

The estimation results are presented in Tables 3 and 4. The dependent variable is the (log of the) number of 2-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.

[12] In addition to the example in the text, the other 4 states where the models differ in their prediction for Coke purchases are when the Coke state is NS and Pepsi state is either NN or SN, or Coke state is SS and Pepsi state is either NN or SN. Symmetric arguments hold for Pepsi.

4.1.1 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 the same fraction of storing consumers for Coke and Pepsi. 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.

Columns 3 and 4 present estimates under perfect foresight (assumption A3) and defining sales based on actual prices: consumers stockpile if prices at t are lower than $t+1$ prices. The next set of columns continue to assume perfect foresight but use a different definition of a sale. Now a sale is defined as any price below 1 dollar: whenever a buyer observes a price below a dollar they purchase for future consumption.[13] In both cases, the definition of a sale is just used to define the state. We do not average prices within a state: actual prices are used.

The final set of columns continue to define a sale as any price below 1 dollar but assume rational expectations instead of perfect foresight. As we discussed in Section 2.1.5, the rational expectations model requires us to either solve a system of equations, or, estimate the model by restricting the sample to periods in which predicted demand does not depend on future prices. This is the approach we follow in the last columns. Since in the states where future price expectations do not matter the perfect foresight and rational expectations models deliver the same predictions the only difference between the estimation in columns 5 and 6 is the periods used. Because of the conditioning, to such states, the number of observations goes down from 45,434 to 30,725.[14]

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 three sets of results is in the cross-price elasticity of storers. The lower cross price effects under perfect foresight perhaps suggests A3 introduces measurement error. Recall the key complication of the rational expectations assumption was future cross prices effects. Eliminating the periods where the assumption

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

[14] The main reason to present the model in columns 5 and 6 is to separately show the effect of the change in the definition of a sale from the effect of the change in the sample.

bites and perhaps fails may raise the cross price estimates. For of the calculations below the difference between the models is not of great importance.

The parameter ω measures the relative intercepts of the demand for 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 be 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.

Table 3: Estimates of the Demand Function

	Static Model		Dynamic Models					
			PF		PF-Alt Sale Def		RE	
	Coke	Pepsi	Coke	Pepsi	Coke	Pepsi	Coke	Pepsi
P_{own} non-storers	-2.30	-2.91	-1.41	-2.11	-1.49	-2.12	-1.27	-1.98
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
P_{cross} non-storers	0.49	0.72	0.61		0.71		0.63	
	(0.01)	(0.01)	(0.01)		(0.01)		(0.01)	
P_{own} storers			-4.37	-5.27	-4.38	-5.08	-5.57	-6.43
			(0.06)	(0.06)	(0.07)	(0.07)	(0.12)	(0.12)
P_{cross} storers			0.61		0.34		2.12	
			(0.04)		(0.04)		(0.09)	
ω	–	–	0.14		0.10		0.21	
			(0.01)		(0.01)		(0.02)	
# of observations			45434		45434		30725	

Note: All estimates are from least squares regressions. The dependent variable is the (log of) quantity of Coke sold at a store in a week. All columns include store fixed effects. Columns labeled PF use perfect foresight (Assumption A3), while columns labeled RE use rational expectations (Assumption A3'). Columns 3 and 4 define the states using actual prices, while the last four columns use the alternative definition of a sale as any price below 1 dollar. The RE model uses the alternative definition of sale but restricts the sample. See the text for details. Standard errors are reported in parentheses.

The own price elasticity implied by the estimates from the dynamic model evaluated at the quantity weighted price for Coke is 2.16 and for Pepsi 2.78, while 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.

4.1.2 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 3-4 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 the storers.

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

In columns 5-6 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% of consumers store for $T = 2$. Thus, the $T = 2$ type are non significant relative to the $T = 1$ type.[15] The last pair of columns paint a similar picture. Here the fraction of consumers who store for 3 periods is less than 5%. These results suggest that $T = 1$ is the preferred option.

Table 4: Estimates of the Demand Function

	PF-T=2		PF-T=3		PF-T=1,2		PF-T=1,3	
	Coke	Pepsi	Coke	Pepsi	Coke	Pepsi	Coke	Pepsi
P_{own} non-storers	-1.62	-2.37	-1.97	-2.42	-1.44	-2.09	-1.43	-2.10
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
P_{cross} non-storers	0.67		0.69		0.64		0.64	
	(0.01)		(0.01)		(0.01)		(0.01)	
P_{own} storers	-4.25	-5.47	-5.31	-5.09	-4.34	-5.20	-4.31	-5.21
	(0.07)	(0.09)	(0.12)	(0.11)	(0.06)	(0.06)	(0.06)	(0.07)
P_{cross} storers	0.81		0.78		0.65		0.79	
	0.04		(0.09)		(0.04)		(0.04)	
ω	0.31		0.41		0.15		0.17	
	(0.02)		(0.03)		(0.01)		(0.12)	
$\omega_{T=1}$					0.84		0.78	
					(0.02)		(0.02)	

Note: All estimates are from least squares regressions. The dependent variable is the (log of) quantity of Coke 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 length of T and the number of different types of consumers. Standard errors are reported in pharenthesis.

[15]Since the price sensitivity of all storers is the same the relative fraction of $T = 1$ consumers and $T = 2$ consumers will stay the same at all prices.

4.2 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 its welfare implications. We consider two benchmarks: (non-feasible) third degree discrimination, targeting storers and non-storers with different prices, and nondiscrimination, a single price for all consumers.

The analysis neglects the vertical relation between manufacturer and retailer, which could generate double marginalization. The relation between retailers and manufacturers is quite interesting and subtle, but beyond the scope of this paper. The first order conditions in equation (8) represent either a manufacturer selling to a competitive retailing industry or integrated pricing with transfers (which avoids double marginalization).

4.2.1 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 get an implied markup for Coke of 43 cents, and 34 cents for Pepsi. Subtracted from a quantity weighted average transaction price of 1.07 and 1.01, respectively, leads 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 an implied margin for Coke is 50 cents and a marginal costs of 57, while 37 and 64 cents for Pepsi. The lower dynamic elasticities translate into higher implied mark-ups.

Naturally, demand dynamics render the static first order conditions inadequate as a description of seller behavior. So we now turn to the dynamic pricing model. We compute prices and profits under non-discrimination, third degree discrimination and sales. We assume 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 regular, non-sale, prices and sale prices for Coke and Pepsi under the different pricing assumptions. We show profits relative to discrimination. By discrimination we mean the case where the firms can identify the storers and non-storers, set different prices for each group and prevent arbitrage. This is of course non-feasible 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 in 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% higher Coke prices than storers' prices. The gap is 33% 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 3-4 of Table 3 ($T = 1$ and perfect foresight) and in row 4 we present, for robustness, the results for the $T = 2$ model with perfect foresight (columns 1-2 in Table 4). In all cases we assume the competitor charges the non-discriminatory price. The exercise amounts to evaluating the impact of different pricing taking as given competitor's behavior. In the next section we evaluate equilibrium regimes where both players discriminate.

The optimal sale price is in 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 for Coke a sale price about 8% below the non-discriminatory price, and 5% 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% of the discriminating profits, while sales accrue about 91% of the discrimination benchmark.

Table 5: Gains from sales

		Coke			Pepsi		
#	Pricing	Regular($)	Sale($)	Profit(%)	Regular($)	Sale($)	Profit (%)
1	Non-discrimination	1.11	–	88	1.04		92
2	Discrimination			100			100
	non-storers	1.31			1.14		
	storers	0.83			0.86		
Sales							
3	T=1	1.31	1.02	91	1.14	0.99	94
4	T=2	1.27	0.95	95	1.14	0.95	97

Note: Computed based on the estimates of columns 3-4 of Table 3, and columns 1-2 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 mc are 0.6 for Coke and 0.67 for Pepsi.

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 oppose to just one period ahead as in the $T = 1$ model). In turn, sales more effectively capture additional gains from discrimination, 95% and 97% of the target for Coke and Pepsi respectively. Thus, capturing around 50% of the gap in profits between discrimination and non-discrimination.

4.2.2 Welfare

The welfare consequences of third degree price discrimination had been studied since Robinson (1933). The standard intuition is that in a monopoly situation price discrimination will yield 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.[16] So while the seller 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 (1984), 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 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 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.

In this section we evaluate the impact of intertemporal discrimination on quantity and welfare. We first consider the implication of our estimates for a monopolist (i.e., a duopolist that unilaterally best responds to 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 these competitors do not price discriminate. This allows us to isolate the impact of different pricing strategies. An additional advantage of this exercise is

[16] 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 (see Nahata, et al. (1990))

that it is linked to the theoretical results in Schmalensee (1981) and Aguirre et. al. (2010), since the seller is basically a monopolist.

It is worth mentioning that the third-degree discrimination theoretical results apply to the intertemporal price discrimination as well. We just need to reinterpret demand during sale periods, $Q^{NS}(p) + 2Q^{S}(p)$, as the weak demand, and demand during non-sale periods, $Q^{NS}(p)$, as the strong demand.

Before looking at the numbers in Table 6 we turn to the theoretical literature for predictions and to make sure the functional forms we use are not responsible for our findings. Proposition 3 in Aguirre et al. (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 et al. call the IRC condition[17] holds for exponential demands, thus Proposition 3 therein applies. Proposition 3 is a comparative static with respect to the degree of price discrimination.[18] 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 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 the 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% of the potential profits from this group for Coke. The overall gains in profits is smaller because sales decrease profits slightly from the non-storing group. 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.

[17]The condition requires the ratio of the derivative of welfare with respect to price to the second derivative of profits be monotonic in price, in each market.

[18]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)

	Coke			Pepsi		
	Price	Quantity	Profit	Price	Quantity	Profits
Non-Discrimination		318.00	161.86		425.27	157.45
non-storers	1.11	258.42	131.54	1.04	345.84	127.96
storers	1.11	59.58	30.32	1.04	79.43	29.39
3rd Degree Discrimination		397.44	184.58		482.79	170.60
non-storers	1.31	194.92	138.20	1.14	277.70	131.63
storers	0.83	202.52	46.38	0.86	205.09	38.97
Intertemporal Discrimination		353.12	167.78		446.06	160.44
non-storers-non-sale	1.31	194.92	138.20	1.14	277.70	131.63
non-storers-sale	0.99	307.36	118.64	0.98	394.18	121.41
storers - non-sale	1.31	0	0	1.14	0	0
storers - sale	0.99	101.98	39.36	0.98	110.12	33.92

Note: Computed based on the estimates of columns 3-4 of Table 3. Each entry shows the price, in dollars, or quantity, in units per week per store, and profits, in dollars per week per store, from each group under each regime.

Equilibrium We 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 feasible).

The first step is to check for quantity increases, absent them, 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, that 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 where 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,[19] relative to the non-discrimination case. In the case of third degree discrimination non-stores are worse off but storers are better off because they are

[19] 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

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.

Total consumer welfare is down under third degree discrimination: the gains to the storers are out weighed 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.

Table 7: Equilibrium Welfare Effects – Coke

	Price	Quantity	Profit	ΔCS	ΔProfits	ΔTS
Non-Discrimination		318.00	161.86			
non-storers	1.11	258.42	131.54	–	–	–
storers	1.11	59.58	30.32	–	–	–
3rd Degree Discrimination		389.15	188.82	-12.33	26.96	14.62
non-storers	1.31	207.69	147.26	-45.04	15.72	-29.32
storers	0.83	181.46	41.56	32.71	11.24	43.95
Intertemporal Discrimination		350.02	168.65	4.54	6.78	11.33
non-storers - non-sale	1.31	207.69	147.26	-45.04	15.67	-29.32
non-storers - sale	0.99	295.96	114.24	34.71	-17.3	17.41
storers - non sale	1.31	0	0	–		–
storers - sale	0.99	98.19	37.90	9.70	7.58	17.28

Note: Computed based on the estimates of column 3 in Table 3. Each entry shows the price, in dollars, quantity, in units per week per store, or profits/changes in welfare, in dollars per week per store, from each group under each regime The last three columns present the change in consumer surplus, profits and total surplus relative to non-discrimination, respectively.

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.

Table 8: Equilibrium Welfare Effects – Pepsi

	Price	Quantity	Profit	ΔCS	ΔProfits	ΔTS
Non-Discrimination		425.27	157.35			
non-storers	1.04	345.84	127.96	–	–	–
storers	1.04	79.43	29.39	–	–	–
3rd Degree Discrimination		486.62	181.56	-16.38	24.21	7.83
non-storers	1.14	313.73	148.71	-36.49	20.75	-15.74
storers	0.86	172.89	32.85	20.10	3.46	23.56
Intertemporal Discrimination		441.87	162.14	-2.22	4.79	2.57
non-storers - non-sale	1.14	313.73	148.71	-36.49	20.75	-15.74
non-storers - sale	0.98	365.68	112.63	21.25	-15.33	5.92
storers - non sale	1.14	0	0	–	–	–
storers - sale	0.98	102.16	31.47	5.40	2.08	7.48

Note: Computed based on the estimates of column 4 in Table 3. Each entry shows the price, in dollars, quantity, in units per week per store, or profits/changes in welfare, in dollars per week per store, from each group under each regime The last three columns present the change in consumer surplus, profits and total surplus relative to non-discrimination, respectively.

5 Concluding Comments

We study the impact of price discrimination when consumers can anticipate demand and 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 can recover 24-30% 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, perhaps over-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 consumer expectations on how prices evolve (see, for example, Erdem, Imai and Keane, 2003, or Hendel and Nevo, 2006b). It depends on the application whether it is more reasonable to make the assumptions of this paper or those of the previous literature. If one is willing to make the assumptions herein the analysis is significantly simpler and in many issues, like the supply side, becomes tractable.

6 References

Aguirre, Inaki, Simon Cowan and John Vickers (2010) "Monopoly Price Discrimination and Demand Curvature" *American Economic Review,* Volume 100, Number 4, September, pp. 1601-1615(15)

Berry, Steven, James Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Blattberg, R. and S. Neslin (1990), Sales Promotions, Prentice Hall.

Boizot, C., J.-M. Robin, M. Visser (2001), "The Demand for Food Products. An Analysis of Interpurchase Times and Purchased Quantities," *Economic Journal*, 111(470), April, 391-419.

Borenstein, Severin (1985). "Price Discrimination in Free-Entry Markets" *The RAND Journal of Economics,* Vol. 16, No. 3 (Autumn, 1985), pp. 380-397.

Chevalier, Judith and Anil Kashyap. (2010) "Best Prices" mimeo University of Chicago.

Conlisk J., E. Gerstner, and J. Sobel "Cyclic Pricing by a Durable Goods Monopolist," *Quarterly Journal of Economics*, 1984.

Corts, Kenneth. (1998) "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment", *The RAND Journal of Economics,* Vol. 29, No. 2, pp. 306-323.

Erdem, T., M. Keane and S. Imai (2003), "Consumer Price and Promotion Expectations: Capturing Consumer Brand and Quantity Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5-64.

Hendel, Igal and Aviv Nevo "Sales and Consumer Inventory," Manuscript, *Rand Journal of Economics*, 2006a.

Hendel, Igal and Aviv Nevo "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 2006b.

Holmes, Thomas (1989) "The Effect of Third Degree Price Discrimination in Oligopoly" *American Economic Review,* Vol. 79, No. 1, March, pp. 244-250

Hong P., P. McAfee, and A. Nayyar "Equilibrium Price Dispersion with Consumer Inventories", *Journal of Economic Theory*, Vol 105, Issue 2, Pages 503-517.

Jeuland, Abel P. and Chakravarthi Narasimhan. “Dealing-Temporary Price Cuts-By Seller as a Buyer Discrimination Mechanism” *Journal of Business*, Vol. 58, No. 3. (Jul., 1985), pp. 295-308.

Lal, Rajiv (1990). "Manufacturer Trage Deals and Price Deal Promotions," *Journal of Marketing Research*, Vol XXVII, 428-44.

Lazarev, John (2012). "The Welfare Effects of Intertemporal Price Discrimination: An Empirical Analysis of Airline Pricing in U.S. Monopoly Markets."

Nahata, B., Ostaszewski, K. and Sahoo, P.K. “Direction of Price Changes in Third-Degree Price Discrimination.” *American Economic Review*, Vol. 80 (1990), pp. 1254-1258.

Perrone, Helena (2010). "Inventories, Unobservable Heterogeneity, and Long-Run Elasticities" mimeo Pompeu Fabra.

Pesendorfer, Martin (2002), “Retail Sales. A Study of Pricing Behavior in Supermarkets,” *Journal of Business*, 75(1), 33-66.

Robinson, Joan (1933), "The Economics of Imperfect Competition" McMillan and Co.

Salop S. and J. E. Stiglitz. (1982) “The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents” *The American Economic Review*, Vol. 72, No. 5, Dec., pp. 1121-1130.

Schmalensee, Richard (1981) "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination" *American Economic Review,* Vol. 71, No. 1 (March), pp. 242-247.

Shepard, Andrea. (1991) "Price Discrimination and Retail Configuration," *The Journal of Political Economy,* Vol. 99, No. 1, Feb., pp. 30-53

Sobel, Joel. “The Timing of Sales,” *Review of Economic Studies*, 1984.

Varian, Hal “A Model of Sales” *The American Economic Review*, Vol. 70, No. 4. (Sep., 1980), pp. 651-659.

Varian, Hal (1985) "Price Discrimination and Social Welfare" *American Economic Review*, Vol. 75, No. 4 September, pp. 870-875.

Verboven, Frank.(1996) "International Price Discrimination in the European Car Market," *The RAND Journal of Economics,* Vol. 27, pp. 240-268

Villas Boas, S. (2009) "An empirical investigation of the welfare effects of banning wholesale price discrimination," *The Rand Journal of Economics,* Vol. 40, pp 20-46.

7 Appendix: 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 equations (3) and (5), modifying for simplicity of presentation the indicator functions in (5) to ignore ties in effective prices , the purchases predicted by the model are given by

$$x_{jst} = q_{jst}^{NS} + x_{jst}^{S} = Q_{jst}^{NS}(p_{jst}, \mathbf{p}_{-j,s,t}) + \sum_{\tau=0}^{T} Q_{jst+\tau}^{S}(p_{jst}, \mathbf{p}_{-j,s,t+\tau}^{ef})\mathbf{1}[p_{jst} = \mathbf{p}_{j,s,t+\tau}^{ef}]$$

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 6). 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 consists 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. Since purchases are scaled differently in different states in order to account for store fixed effects we need to transform the predicted purchases as follows. Given the functional form in equation (10) and assuming the fraction of non-storers is given by ω,

$$Q_{jst}^{NS}(p_{jst}, \mathbf{p}_{-j,s,t}) = \omega e^{\alpha_{sj}} e^{\{-\beta_j^{NS} p_{jst} + \gamma_{ji}^{NS} p_{ist}\}} e^{\varepsilon_{jst}}$$

$$Q_{jst+\tau}^{S}(p_{jst}, \mathbf{p}_{-j,s,t+\tau}^{ef}) = (1-\omega) e^{\alpha_{sj}} e^{\{-\beta_j^{S} p_{jst} + \gamma_{ji}^{S} p_{ist+\tau}^{ef}\}} e^{\varepsilon_{jst+\tau}}$$

Denote by

$$Q_{jst}^{*NS} = \omega e^{\{-\beta_j^{NS} p_{jst} + \gamma_{ji}^{NS} p_{ist}\}} e^{\varepsilon_{jst}} \quad \text{and} \quad Q_{jst+\tau}^{*S} = (1-\omega) e^{\{-\beta_j^{S} p_{jst} + \gamma_{ji}^{S} p_{ist+\tau}^{ef}\}} e^{\varepsilon_{jst+\tau}}$$

then

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

and

$$\log(x_{jst}) - \overline{\log(x_{jst})} =$$
$$\log\left(Q_{jst}^{*NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{*S}(p_{jst}, \mathbf{p}_{-j,s,t+\tau}^{ef}) \right) - \overline{\log\left(Q_{jst}^{*NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{*S}(p_{jst}, \mathbf{p}_{-j,s,t+\tau}^{ef}) \right)}$$

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

$$\overline{\log\left(Q_{jst}^{*NS} + \sum_{\tau=0}^{T} Q_{jst+\tau}^{*S}(p_{jst}, \mathbf{p}_{-j,s,t+\tau}^{ef}) \right)}$$

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.