INTERTEMPORAL SUBSTITUTION AND STORABLE PRODUCTS

Igal Hendel
University of Wisconsin, Madison
and NBER

Aviv Nevo
University of California, Berkeley
and NBER

Abstract
Storable products allow consumers to time their purchases to exploit price fluctuations. It has been documented that during promotions consumers buy more. The additional purchases are potentially intended not only for current use, but to be stockpiled for future consumption. This paper discusses the predictions of a consumer inventory model and reviews the available evidence. We then discuss the implications for demand estimation and present estimates of the economic magnitude of the dynamic effect of storability. (JEL: L0, L4, D1, D4)

1. Introduction

Consumers’ ability to inter-temporally substitute consumption, or at least purchases, has far reaching implications for demand estimation. When consumers can time their purchases demand becomes a function not only of current prices, but also depends on expected future prices. For example consider the replacement of a durable product, like a car. The replacement decision is contingent on current as well as quality-adjusted expected future prices and the current state of the car. In order to consistently estimate the demand function we need to take account of these additional factors, and in particular model consumer expectations about future prices. Similar issues arise in modeling demand for non-durable storable products. Storable products provide consumers with the ability to time their purchases to exploit price fluctuations.

In this paper we review the economics and marketing literature that focuses on non-durable storable products. We summarize the evidence suggesting that intertemporal substitution might be important, we discuss the implications for the interpretation of demand elasticities and we offer estimates of the magnitude of the effect of storability. Like most of the literature we review, we focus on products sold in supermarkets and observed in scanner data. These products are

Acknowledgments: We gratefully acknowledge support from the NSF and the Food System Research Group at University of Wisconsin.
E-mail addresses: Hendel: igal@ssc.wisc.edu; Nevo: nevo@econ.berkeley.edu.
purchased at a high frequency, have frequent price reductions and are storable at some cost. While the price reductions are not perfectly predictable, it is much easier for buyers to realize that an item is on sale than judging whether prevailing interest rates call for an immediate replacement of a durable.

The importance of properly estimating demand when goods are storable stems from the distinction between reactions to short run and long run price changes. Price reductions may have two effects on quantity bought: first, a consumption effect if consumption is price sensitive, second, a stockpiling effect if dynamic considerations lead consumers to accumulate inventory for future consumption. For example, in the sample used in Hendel and Nevo (2002a) the quantity of laundry detergents sold is 4.7 times higher during sales than during nonsale periods (provided there was no sale the previous week). Instead if there was a sale in the previous week, then the quantity sold is only 2.0 times higher. This pattern suggests not only that demand increases during sales, but that demand accumulates between sales.

This pattern of sales also suggests that standard, static, demand estimation may provide misleading estimates. When the data available for demand estimation presents frequent price reductions (as is the case with scanner data) standard static demand estimation would capture short run reactions to prices, which reflect both the consumption and stockpiling effects. In contrast, for most demand applications (e.g., merger analysis or computation of welfare gains from introduction of new goods) we want to measure long run responses.

In order to test the intertemporal substitution the literature has focused on a consumer inventory model. In the model consumers balance the cost of holding additional inventory with the potential gains of buying at a low price (relative to expected future prices). This model, in contrast to the typical sS model of durable-goods purchases, predicts that the demand acceleration is to be followed by a longer duration to the next purchase. In contrast, the acceleration of the purchase of a durable, induced by low prices, in turn accelerates the subsequent replacement of that good. The nature of the intertemporal substitution of storable good purchases is different from other products, and its proper modeling is needed to estimate demand.

The main hurdle in documenting demand patterns of storable goods is that inventories are unobserved. If we observed the consumers' inventories then determining whether consumers stockpile in response to price movements would be straightforward. For instance, we could test if after a sale end-of-period inventories are higher. However, consumption and therefore inventory, are unobservable. Researchers have taken different approaches to handle the

1. Another source of inter-temporal substitution in the absence of product storability and durability is non time separability of preferences. See Hartmann (2003) who studies intertemporal effects in the demand for recreational golf.
lack of data on inventories ranging from imputing inventories in a structural way to testing predictions that indirectly testify on inventory behavior.

In the rest of this paper we describe the theoretical predictions put to test, the typical data used, survey the main findings, and discuss the implications of the findings for demand estimation.

2. Theoretical Implications of Stockpiling

The products studied are usually sold at a “regular” price with occasional temporary price reductions. Storable products enable the buyer to balance the benefits from buying at the low price with the cost of holding the inventory. For a formal model see, for example, Arrow, Harris, and Marschak (1951), Blattberg, Eppen, and Lieberman (1981), or Boizot, Robin, and Visser (2001).

There are several implications of a simple inventory model which have been tested. First, the quantity purchased is a decreasing function of price. The standard neo-classical static economic model will also predict this effect: if price goes down, consumers consume more. Here the effect exists even if one believes that consumption does not respond at all to prices. To separate the two theories one can look at the effect of past prices on current demand. In the inventory model quantity sold is a function of current inventory, or in lieu of information on current inventory, demand should be a function of past prices (more precisely, the price during the last purchase).

Second, the timing of the purchase is affected by current and previous prices. When buying on sale, the duration to previous purchase should be shorter and the duration to next purchase should be longer (compared to a purchase during a nonsale period).

Third, both the likelihood of purchasing and the amount purchased conditional on a purchase should be declining in the inventory already held. This is driven by the increasing (and convex) costs of holding inventory. The higher the current inventory the higher the cost of storing more and therefore the consumers are less likely to buy, and likely to buy less if they buy.

Fourth, aggregating consumer-level behavior the quantity sold at the store level should increase during a sale, both because consumers are buying more and because they are buying earlier. Moreover, quantity sold should depend on the duration since the previous sale. The longer it has been since the previous sale the lower on average are the inventories held by consumers. Finally, due to the same logic the quantity sold during nonsale periods increases in the duration from the last sale.

Several studies had paid attention to retailer pricing of storables. Salop and Stiglitz (1982) presented the first model where consumers can store a unit for future consumption. Storability creates equilibrium price dispersion. Hong,

3. Typical Data

The typical data used comes from supermarkets. Two companies, AC Nielsen or IRI, buy the data from the supermarkets and sell it to manufacturers. The data typically has two components, store and household-level.

The store level data are collected using scanning devices in supermarkets. For each detailed product (defined by the bar code on the label) in each store in each week the data contains information on the price charged, (aggregate) quantity sold, and promotional activities that took place. The data also generally includes a description of each product, including size and brand information, useful to figure out categories of products.

An additional component of the data set, which is not present in all data sets, is at the household-level. Households who agree to participate in the sample are generally tracked for a period of a year or two. During this period an effort is made to ensure that all the purchases of the household are measured. For each purchase it is known the exact product purchased and the number of units purchased. In addition the data contains detailed demographics of the household.

4. Do Consumers Stockpile?

The interest in consumers’ reactions to promotions, particularly the potential of stockpiling during sales, has long been present in the marketing literature. With the increasing use of scanner data in economics, the interest in stockpiling has risen in the economics literature as well. Several approaches had been used in both fields to document the dynamics generated by storability. Here we survey the different approaches and summarize the evidence suggesting that consumers stockpile.

4.1. Duration and Quantity Effects: Household Data

The marketing literature has concentrated mostly on two predictions of the inventory model, both tested with household data. These are: an increase in the
quantity purchased during a sale, and sales lead to changes in the timing of purchases.\footnote{For example, see Ward and Davis (1978), Shoemaker (1979), Wilson, Newman, and Hastak (1979), Blattberg, Eppen, and Lieberman (1981), Neslin, Henderson, and Quelch (1985), Moriarty (1985), Gupta (1988), Chiang (1991), Grover and Srinivasan (1992), and Bell, Chiang, and Padmanabhan (1999).}

Blattberg, Eppen, and Lieberman (1981) report that in the four categories they study there is an increase in duration to next purchase ranging from 23 to 36\% and an increase in quantity purchased from 8 to 35\%. Further work found results similar in magnitude. For a more comprehensive review of the marketing literature and the findings see Blattberg and Neslin (1990).

Boizot, Robin, and Visser (2001) tests whether duration from previous purchase increases in current price and declines in past price, and quantity purchased increases in past prices. They use a survey of french households. It is similar to the scanner data but limited in terms product information: households report the purchase of an item, but the brand and container size are not known. Their findings are mostly in support of an inventory model, both in terms of quantity and duration effects. Hendel and Nevo (2002a) report both quantity and duration effects in a sample of soda, detergent, and yogurt purchases. Evidence of timing effects of sales is supported by positive duration effects forward, and negative effects in duration backwards (i.e., to previous purchase).

The early marketing literature generally found much larger effects than the economics papers just described. The discrepancy seems to be an artifact of the estimation used in the marketing literature. The data used in the analysis are panel data: purchases of different households are observed over time. Therefore, there are different estimates of the effects. Consider for example the effect on quantity purchased. The Total difference in quantity is the difference between the average, across households and time, quantity sold during sales and the average during nonsales. The Within effect allows for differences across households in the typical quantity purchased. It is computed by taking the difference in the average quantity purchased by each household during sale and nonsale periods, and then averaging the effects across the households. The two estimates will be different when the typical quantity purchased by a household is correlated with the households tendency to buy on sale (e.g., larger households buy more on sale). The Between effect measures differences across households in their typical, or average, behavior.

With the exception of Shoemaker (1979),\footnote{While Shoemaker does not call his estimate a Within estimate, his procedure is identical to what we call Within.} the literature that tested the inventory model’s predictions at the household level did so by comparing the Total difference across sales and nonsales periods. Instead, to test the inventory models one wants to look at Within household implications (Hendel and Nevo...
Consequently, most of the findings in the literature confound two distinct effects, one of which, a Between difference in sale and nonsale purchases, should be purged in order to quantify the effect of promotions. The household reactions to sales have been magnified by the between effect, contributing to the postpromotion dip puzzle (see Section 4.3).

The reason between household differences magnify the sale/nonsale differences is that on average households who tend to buy more on sale also tend to buy less frequently. This cross-household pattern does not tell us whether the average household is buying earlier because of the promotion. It tells us that buyers are heterogeneous. This heterogeneity creates a large total effect due to between differences, but not due to promotions’ effect within household. Once the Between difference are purged the sales versus nonsales differences are statistically significant but are relatively small, compared to the Total difference. They predict roughly a change of 5–10% in the interpurchase duration (Hendel and Nevo 2003a), more modest than the 8–36% reported previously. Nevertheless, the effects are still both statistically and economically significant.

4.2. Implied Inventories: Household Data

An alternative approach is based on generating a proxy for inventories and using it to test the effect of inventory on the decision to purchase and the quantity purchased conditional on purchase. Neslin, Henderson, and Quelch (1985) use this approach to study the reaction of purchases, for coffee and bathroom tissue, to various promotional activities (including advertised price cuts). They impute inventory by assuming that weekly consumption, for each household, is constant over time and can be computed from the total purchases divided by the number of weeks; and initial inventory is equal to the average purchased amount. Using these assumptions and observed purchases they can impute inventory over time. To test the impact of promotions on quantity purchased and time since the previous purchase they regress these variables on the imputed inventory and dummy variables for different promotional activities. They find substantial effects of promotions on duration to next purchase as well as on quantity purchased. Subsequent work (Gupta, 1988, 1991; Bucklin and Lattin 1991; and Currim and Schneider 1991) also finds significant effects of inventory.

Neslin and Schneider Stone (1996) repeated this exercise and also found statistically significant effects. They used the estimated coefficients to simulate the likely effect of inventory behavior on aggregate demand. They claim that the effects are small. Indeed, they offer these as an explanation to the so-called

---

4. Alternative approaches to imputing the initial inventory have also been employed in the literature. These include: setting the initial inventory to a level just high enough so the imputed inventory will not be negative or by including a household fixed-effect that captures the effect of the initial inventory.
postpromotion dip puzzle (see Section 4.3). Hendel and Nevo (2002a) also
found a statistically significant but economically small effect of inventory on
probability of purchase and on quantity purchased. They attribute it to econo-
metric issues, such as measurement error in inventory.

4.3. Store-Level Data and the Postpromotion Dip Puzzle

The papers described in Sections 4.1 and 4.2 found evidence at the household-
level consistent with the stockpiling model. As we discussed in Section 2, the
model implies that aggregate quantity sold both during sale and nonsale periods
should be a function of the duration from the previous sale, or more generally
any promotional activity. Several papers study these predictions.

Pesendorfer (2002) uses store level data of ketchup sales to study the effect
of the duration since the last sale on the probability to hold a sale and on the
quantity sold. He proposes a model in which a fixed number of consumers
appears every period. These consumers differ in the willingness to pay (as in the
Sobel 1984) and in their store loyalty. He shows that in equilibrium the decision
to hold a sale is a function of the duration since the last sale. His empirical
analysis shows that both the probability of holding a sale and the aggregate
quantity sold (during a sale) are a function of the duration since the last sale. The
latter is evidence of demand accumulation, which he interprets as support for a
Sobel-type model where consumers accumulate in the market until they buy. Although he models demand à la Sobel, his numbers testify on stockpiling. He
shows that demand during sales increases in the duration since last sale.

Another prediction of an inventory model, that helps distinguish it from a
Sobel type model, is that the quantity sold during non-sale periods should also
increase in the duration since last sale. In particular the model predicts a dip in
the quantity sold after a sale. This dip has been hard to consistently find
(Blattberg and Neslin 1989). The failure to find a dip has been named the
postpromotion dip puzzle. Several explanations have been offered for the lack
of a postpromotion dip. Neslin and Schneider Stone (1996) discuss eight
possible arguments to sort out the apparent difference between the household-
level data and the aggregate data. They seem to favor an explanation that the
household effects—as measured by the effect of an imputed inventory on
purchase decisions (see discussion in Section 4.2)—is small. Van Heerde,
Leeflang, and Wittink (2000) take a different approach to studying the puzzle.
They propose a distributed lag analyses of aggregate weekly sales data. Indeed
their results show that adding various lags and leads can help to find the dip.

Hendel and Nevo (2003a) combine these two explanations. For the house-
hold data they note that in order to test the model one has to separate the Total
effect from a Within household effect (see the discussion in Section 4.1).
Furthermore, they show that in order to find the desired effect in the aggregate
data one needs to control for dynamic effects of additional promotional activity, which are correlated with price reductions. Once they regress the total quantity sold on duration from previous sale, duration from previous promotions, current prices, advertising, and various fixed effects, they find the desired effects both during sales and non-sale periods. Their results suggest that the expected dip is present for sales that do not have the additional promotions. While promoted sales will not have the expected effect because of a counter dynamic effect of the promotions.

4.4. Cross Category Comparisons

Boizot, Robin, and Visser (2001) and Hendel and Nevo (2002a) compare categories with different degree of storability. Boizot, Robin, and Visser distinguish storable products, like rice and sugar, from partially storable products like fruits. Interestingly, results for the partially storable products are less aligned with the predictions of the model. More precisely, the dependence of quantity sold on past prices is significant and positive for storables, but mostly independent of past prices for less storable products.

Hendel and Nevo (2002a) compared the results, for several storable products, to those obtained for milk. The retail price of milk exhibits a very different pattern than the one shown by storable categories. The main difference is the absence of temporary price reductions. Assuming sales are motivated by a desire to discriminate across consumers with different ability to store, there should be no sales for milk. Indeed, this seems to be the case.

Another cross-category comparison discussed in Hendel and Nevo (2002a) involves the difference between laundry detergents and yogurt. Since the average duration between supermarkets visits is less than a week, both these products are storable. However, there is a key difference between how one would store them. Unlike detergents, the storability of yogurt decreases once a container is opened. This suggests that for detergents we should see more sales for larger sizes and when consumers purchase on sales they buy larger units. For yogurt we should see the opposite: more sales for smaller containers and purchase of more units of smaller sizes during sales. Both these predictions are found in the data.

5. Demand Estimation

The findings reported above suggest that dynamic considerations impact consumers purchasing decisions. Failing to control for these dynamic consider-
ations is likely to have undesirable consequences in demand estimation. Standard demand models that neglect inventories will be mis-specified. Even if we properly control for inventories, neglecting the storage element of demand provides estimates of short run price elasticities. Short run elasticity estimates are likely to overstate consumers’ long-run price responses, which involve consumption responses but no stockpiling. Static demand estimates instead are likely to capture both the consumption and stockpiling effects.

Hendel and Nevo (2002a) present a simple exercise, to approximates the overstatement of long run estimates if one neglects dynamics. They proxy the short-run elasticities by the increase in demand associated with a sale. On the other hand, they proxy the long-run elasticity using interpurchase duration data to purge the elasticities from the stockpiling effect. The idea is to compute a duration-corrected change in demand, which is done by spreading the purchases after a sale, over the longer durations to next purchase that take place after a sale. The results suggest that neglecting dynamics leads to an overstatement of demand elasticity by a factor of 2 to 6, for the products they study.

Erdem, Imai, and Keane (2003) and Hendel and Nevo (2002b) take a structural approach to assess the distinction between short run and long run price responses, and to assess the impact of neglecting demand dynamics. Both papers use household-level data to estimate dynamic discrete choice models of demand in which consumers can store different varieties of the product. The papers differ in how they handle the computational complexity of the dynamic decision problem in light of the large choice set faced by households. The advantage of the structural approach is that enables a better handle of product differentiation, which the simple exercise in Hendel and Nevo (2002a) neglects, and enable researchers to evaluate different policy experiments.

The focus of Erdem, Imai, and Keane (2003) is on simulating consumer responses to short-run and long-run price changes. In contrast, Hendel and Nevo (2002b) compare long run elasticities to those obtained from standard static methods. The latter is the relevant measure of how far static estimates may be from long-run elasticities.

Erdem et al. report, for ketchup, an own price elasticity with respect to a short-run price reduction 29% larger than the long run price elasticity. They also report lower short-run cross price elasticities than long run ones. Hendel and Nevo (2002b) report that static demand estimates, which neglect dynamics, may overestimate own price elasticities by 30%; underestimate cross-price elasticities to other products by up to a factor of 4; and overestimate the substitution to the no-purchase, or outside option, by up to 150%.

Estimates of the demand elasticities are typically used in a first order condition, for example, from a Bertrand pricing game, in order to compute price cost margins (PCM). For single-product firms it is straightforward to see the magnitude of the bias: It is the same as the ratio of the own-price elasticities.
The above numbers suggest that for single-product firms the PCM computed from the dynamic estimates will be roughly 30% higher than those computed from static estimates. The bias is even larger for multiproduct firms since the dynamic model finds that the products are closer substitutes (and therefore a multiproduct firm would want to raise their prices even further). Another use of demand estimates is for simulation of the effects of mergers. The above estimates suggest that the static model would tend to underestimate the effects of a merger, because it tends to underestimate the substitution among products and will favor approval of mergers.

6. Concluding Comments

In this paper we survey the evidence suggesting that consumers stockpile. Several bits of evidence suggest that dynamic are introduced by product stor-ability. We discuss the implications for interpretation of demand elasticities and we offer estimates of the magnitude of the effect. The magnitude of the effects have significant implications for both public policy and optimal firm behavior.

This suggests several directions for future research. First, the structural estimates of the magnitude of the effects reported by Erdem, Imai, and Keane (2003) and Hendel and Nevo (2002a) required several strong assumptions. Future work should try to relax some of these assumptions and apply the methods to additional products. Even with the strong assumptions the methods offered by these papers are computationally intense. Thus, in order to be useful for policy work—and as we saw the magnitude seems to suggest important implications for policy—some approximations would be needed.

Second, given the evidence in favor of stockpiling obvious questions arise about the optimal behavior of firms selling storable products. What is the optimal timing of a sale? What is the optimal discount? How do these change with competition? Since firms might be able to exploit stockpiling by offering larger sizes (at a per unit discount), how should sales and nonlinear pricing interact?

References


