Are Sale Signs Less Effective When More Products Have Them?

Eric T. Anderson • Duncan I. Simester
GSB, The University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637
Sloan School of Management, MIT, E56-305, 38 Memorial Drive, Cambridge, Massachusetts 02139
Eric.Anderson@gsbpop.uchicago.edu • simester@mit.edu

Abstract

We analyze data from a variety of sources, including historical data from a women's clothing catalog, a field study in that catalog, survey responses to catalog stimuli, and grocery store data for frozen juice, toothpaste, and tuna. The analysis yields three conclusions. First, sale signs are less effective at increasing demand when more items have them. Second, total category sales are maximized when some but not all products have sale signs. Third, placing a sale sign on a product reduces the perceived likelihood that the product will be available at a lower price in the future, but the effect is smaller when more products have sale signs. By ruling out alternative hypotheses, the findings suggest that moderation of the sale sign effect is in part due to reduced credibility when they are used on more products.

The credibility argument is motivated in part by a recent paper that presented an equilibrium model predicting that customers who lack knowledge of market prices rely on point-of-purchase sale signs to help evaluate posted prices (Anderson and Simester 1998). The model predicts that sale signs increase demand but that the increase is smaller when more products have them. This moderating effect regulates how many sale signs stores use and makes customer reliance on these cues an equilibrium strategy.

To evaluate whether the number of sale signs moderates their effectiveness we compare demand for items with sale signs when varying the number of sale signs on other products. For this comparison we use two datasets describing demand for products in a women's clothing catalog. The first dataset describes customer orders for the same set of items across three sequential issues of the catalog. The second dataset is a field test conducted by mailing different versions of the catalog to randomly selected customer samples. Reassuringly, although the datasets do not share the same limitations, the findings are consistent and indicate that sale signs are less effective when more products have them.

In addition to the credibility explanation there are at least two alternative explanations for this result. First, using a sale sign to make a product more attractive may lead to substitution of demand from other sale items within a store, and so demand may vary even if there is no change in the credibility of the sale signs. To discriminate between the substitution and credibility explanations, we evaluate whether total category demand is maximized when some but not all products in the category have sale signs. The credibility explanation implies that adding another sale sign will eventually decrease total category sales when many items already have sale signs. In contrast, if adding a sale sign to an item leads solely to substitution from other products within (or outside) the category we will not observe a decrease in total category demand. We test the total category sales prediction using grocery store scan data describing total category demand for frozen juice, toothpaste, and canned tuna. The findings reveal that in all three categories there is a significant reduction in aggregate demand.

A second alternative explanation is that sale signs focus customer attention on products with these cues. Distinguishing between the attention and credibility explanations is a difficult task. They both imply that sale signs deliver less information when there are too many sale signs. The credibility argument predicts that this occurs because the signs are noticed but not believed. The attention explanation predicts that the signs are less likely to be noticed when attention is diluted by a large number of sale signs. To discriminate between the credibility explanation and this attention effect we use survey measures to evaluate a third hypothesis. This hypothesis predicts both that placing a sale sign on a product reduces the perceived likelihood that the product will be available at a lower price in the next period, and that this effect is smaller when more items have them. In an attempt to control for the attention effect we focus subjects' attention on a set of focal items. Under these conditions it is unlikely that subjects overlook sale signs on these items. More important, the likelihood of overlooking sale signs on the focal items is unlikely to depend on how many other items also have sale signs. The results confirm both that the presence of a sale sign reduced subjects' expectations that an item would be available at a lower price in the future and that this effect is smaller when more items have sale signs.

(*Sale Signs; Retail Pricing; Promotions; Credibility; Signaling; Fashion Products*)

0732-2399/01/2002/0121/$05.00
1526-548X electronic ISSN

1. Introduction

Anderson and Simester (1998) recently presented an equilibrium model predicting that customers who lack knowledge of market prices rely upon point-of-purchase sale signs to help evaluate posted prices. The findings confirm that customer reliance on sale signs can be an equilibrium strategy, and illustrate how firms respond to this. The model predicts that sale signs increase demand but that the increase is smaller when more products have them. It is this moderating effect that regulates retailers' use of sale signs and makes customer reliance on these cues an equilibrium strategy. This paper has two goals. First, we present evidence from several sources confirming that sale signs are less effective when more products have them. We then investigate whether this finding can be explained, at least in part, by Anderson's and Simester's (1998) credibility argument. In doing so, we acknowledge that there may exist other circumstances in which alternative explanations lead to similar findings.

The Anderson and Simester (1998) Model

The finding that placing sale signs on a product can increase demand for that product is already well established. In a laboratory experiment, Inman et al. (1990) show that placing a sale sign on an item is sufficient to increase demand for the item without changing the actual price. Inman and McAlister (1993) then show that the result survives in a campus grocery store using promotional cues combined with negligible discounts. Further evidence can be found in discrete choice models. After separately controlling for the regular price and any price discount, there is strong evidence that customers are more likely to purchase items that are accompanied by sale signs (see, for example, Guadagni and Little 1983 and Grover and Srinivasan 1989). The apparent effectiveness of sale signs is surprising; they are inexpensive to produce, and stores generally make no commitment when using them. As a result, they can be placed on any products, and as many products, as stores prefer.

Anderson and Simester's explanation for why sale signs are effective begins by recognizing that many customers lack information about market prices. Uncertainty about future discounts and prices at competing retailers prompt these customers to look for additional cues to help evaluate whether to make an immediate purchase. Retailers respond by placing sale signs at the point of purchase to identify which items are discounted. Customer reliance on these cues may then explain why sale signs are effective at increasing demand. Anderson and Simester argue that customers implicitly regulate the temptation to place sale signs on every item by ascribing less *credibility* to sale signs when they see them on more products. This moderating effect, which has not previously been tested in the literature, leads to the following prediction:

HYPOTHESIS 1. *Sale signs are less effective at increasing demand when more products have them.*

This prediction is an endogenous outcome of the model. Customers use Bayes' rule to update their beliefs, which influence retailer strategies. These strategies are, in turn, consistent with customer beliefs in equilibrium. In particular, customers have prior beliefs about the number of items that are truly discounted. If the number of sale signs exceeds customers' expectations, they infer that not all of the items with sale signs are truly discounted. This loss of credibility makes sale signs less effective at convincing customers that they should purchase immediately rather than visit another store or wait for a future discount.

Alternative Explanations

We test Hypothesis 1 by observing whether the demand for items with sale signs is smaller when more items have them. We caution that there are at least two alternative explanations for this result. First, using a sale sign to make a product more attractive may lead to substitution of demand from other sale items within a store, and so demand for these other sale items may decrease even if there is no change in the credibility of the sale signs. In support of this *substitution* explanation, Gupta (1988) presents evidence that more than 84% of the sales increase due to promotion comes from brand switching. To discriminate between the substitution and credibility explanations, we introduce a second prediction:

HYPOTHESIS 2. *Total category demand is maximized when some but not all products in the category have sale signs.*

Hypothesis 2 implies that total category demand will eventually decrease when more items within the category have sale signs. The benefit from placing a sale sign on an additional item is eventually outweighed by the loss of credibility among the existing sale signs. The resulting concavity regulates how many sale signs a store uses. We caution that Hypothesis 2 focuses on the role of sale signs, not prices, and so we cannot conclude that total category sales ever decrease when adding a sale sign together with a price reduction. Total category sales may continue to expand due to the simultaneous price reduction. For this reason, when testing Hypothesis 2 we explicitly control for price differences.

The prediction assumes that varying the number of sale signs within a category is sufficient to influence the credibility of these sale signs, yet the theory is silent about the domain over which customers judge credibility. If varying the number of sale signs within a category is not sufficient to influence the credibility of these sale signs, the hypothesis will not hold. The risk of rejection is further increased by the possibility of substitution between categories. While total category demand provides an effective control for substitution within the category, it does not control for substitution between categories. If using more sale signs within the category leads to substitution from outside the category, we would expect category expansion rather than contraction. Note that any evidence supporting the hypothesis occurs despite, and not because of, these difficulties.

A second alternative explanation is that sale signs focus customer attention on products with these cues. This effect, which we label an *attention* effect, could lead to sale signs losing their effectiveness if attention is diluted when more items have sale signs (Hypothesis 1). It could also lead to a reduction in total category demand when more items have sale signs if a threshold level of attention is required to prompt purchase (Hypothesis 2). We introduce a third prediction to discriminate between the credibility explanation and this attention effect. This hypothesis summarizes customer beliefs underlying the credibility argument.

HYPOTHESIS 3. *Placing a sale sign on a product reduces the perceived likelihood that the product will be available at a lower price in the next period, but the effect is smaller when more products have them.*

Note first that *substitution* makes no predictions regarding customers' expectations or how a sale sign's effect on expectations will vary according to the number of products with sale signs. With respect to *attention,* a search of the literature reveals no support for a conclusion that focusing attention on an item will lead to expectations that the product is less likely to be available at a lower price in the future. However, it is possible that customers are less likely to attend to a sale sign on an item when their attention is diluted by a large number of sale signs. To address this possibility, we test Hypothesis 3 under conditions that focus subjects' attention on a set of focal items. Under these conditions, evidence supporting Hypothesis 3 is consistent with the credibility argument, but is unlikely to arise due to customers overlooking the sale signs on these focal items.

Empirical Tests

The literature offers little evidence supporting or rejecting predictions Hypothesis 1, Hypothesis 2, and Hypothesis 3. We test these predictions using data from several sources, beginning in §2 and 3 with data from a women's clothing catalog. In §2 we present analysis of historical sales data, followed in §3 with a field test, conducted by mailing different versions of a catalog to randomly selected customer samples. These two datasets allow us to evaluate whether the increase in demand when placing a sale sign on a product (if any) is smaller when more products have sale signs (Hypothesis 1). The field data and historical sales data have different limitations. The cost of undertaking studies in the field limits the extent to which we can manipulate the experimental design. However, by varying the use of sale signs on the same product we overcome the endogeneity limitations inherent in historical data. In comparison, the analysis

Table 1 Summary of Predictions, Theories, and Data

	Predictions from Each Theory			
	Credi-bility	Attention	Substi-tution	Data Used to Test Each Prediction
Hypothesis 1	Yes	Yes	Yes	Catalog field study Historical catalog data
Hypothesis 2	Yes	Yes	No	Grocery store data (3 categories)
Hypothesis 3	Yes	No	No	Survey data

of historical catalog data allows us to generalize the findings to different product categories and to include explicit controls for the degree of substitutability between products.

In §4 we use three grocery store datasets to test the prediction that aggregate demand is maximized when some but not all products have sale signs (Hypothesis 2). The analysis enables us to evaluate whether the findings from the catalog data hold in markets other than fashion clothing and helps us to discriminate between the competing explanations for Hypothesis 1. Finally, in §5 we use survey measures to evaluate Hypothesis 3 by determining whether sale signs affect expectations about future availability and prices and whether this effect (if any) is smaller when more products have sale signs. In Table 1 we summarize the relationship between the three predictions and the alternative explanations and the data used to test each prediction.

This combination of hypotheses and data affords a test of whether sale signs are less effective when more products have them and a means to discriminate between alternative hypotheses for this result. The findings also have broader implications for the estimation of choice models, the selection of retail promotion strategies, and the testing of signaling models. We discuss these and other implications in §6.

Empirical Challenges

Several factors make the task of testing these predictions challenging. First, as noted, the theory is silent as to the domain over which customers judge credibility. In a mail-order catalog, the credibility of sale signs may depend upon the number of sale signs on an individual page, the number on a facing two-page spread, the number within a single product category, or even the number in the entire catalog. In a grocery store it may depend on the number of facings or the number of items within or across product categories. The relevant domain is an empirical question that we address by considering different domains. If the necessary conditions under which the predictions hold are not satisfied, the predictions will be rejected even if they are accurate in other settings.

Second, the theory is also silent as to the specific cues that customers rely on. Inman et al. (1990) introduce the term *promotion signal* to describe any sign, marker, or other indicator of a price promotion, and recognize that such signals may take on a variety of forms. We retain the term *sale sign* to maintain consistency with the Anderson and Simester terminology but recognize that their findings may apply to any of the promotion signals envisaged by Inman et al. (1990). Indeed, the data include examples of several different types of promotion signals. For example, in the women's clothing catalog, promotional prices are identified variously by the words ''Sale,'' ''Pre-Season Sale,'' and ''Clearance.'' We might reasonably expect the effectiveness of these cues to vary.

Third, the theory offers little guidance as to whether customers are sensitive to the number or percentage of items with sale signs. In the historical catalog data we use the percentage rather than the number because the number of items with sale signs is correlated with the number of items on the page, introducing a potential confound (see later discussion). However, in the grocery store data we report the findings for both the number and the percentage. Finally, it is also likely that sale signs perform more than one role. For example, stores may use sale signs to influence customers' expectations about the overall price image of the store and/or to signal which customers a product is targeted at. The possibility that sale signs are serving an alternative role further increases the risk of false rejection.

2. Historical Catalog Data

We received data describing customer orders for three sequential issues of a mail-order catalog selling women's clothing. For confidentiality reasons we are not able to identify the name of the catalog, nor can we describe its target market in detail. The three issues were distributed nationally, approximately 1 month apart, to over 600,000 addresses, including both new and existing customers (although the distribution of the issues overlapped, they were not identical). They contained a mixture of clearance items, sale items, and nonsale items, with the clearance items positioned together in a clearly identified section. Price descriptions for these clearance items included both the regular and discounted price: "Reg $__ SALE $__." Other sale items were dispersed throughout the catalog and included the same price descriptions as the clearance items, "Reg $__ SALE $__." Prices of the nonsale items were simply presented as: "$__."

On the advice of the catalog managers, we omitted any items that had a clearance cue in any issue, as sales of these items are likely to have been truncated by the unavailability of inventory (stock outages for clearance items are common). The main (nonclearance) sections of the first two issues were identical, and contained 220 items, 20 of which were sale items. The main section of the third issue contained 243 items, including 59 sale items. A total of 162 items appeared in the main sections of all three issues. Among these 162 items, 13 had sale signs in all three issues, 1 had a sale sign only in the first and second issues, and 28 had sale signs only in the third issue. As we discuss, the analysis controls for issue effects, so that the tendency for more sale signs to appear in the third issue does not bias the results, even if demand is higher in that issue due to seasonal or distribution differences.

Comparing the prices of the 162 items that appeared in all three issues confirmed that sale signs in this catalog are informative, revealing which items were less expensive in previous issues. Prices of the 41 items with sale signs in the third issue were on average 15% less expensive in that issue than in the first two issues. In contrast, the 125 items without sale signs in the third issue were on average slightly more expensive in that issue (the difference in these price comparisons is significant, $p < 0.01$). Further inspection reveals that the sale signs are not always accurate. Although all of the items that were less expensive in the third issue had sale signs in that issue, not all of the items with sale signs in the third issue were less expensive than in the previous two issues. Eight of the 41 items with sale signs in the third issue had the same price as in the two earlier issues, and 2 of the items with sale signs in the third issue were actually more expensive than in the two previous issues. Anderson and Simester (1998) predict that sale signs offer a noisy signal of which items are discounted, with retailers preferring to first place sale signs on items that are truly discounted but also placing sale signs on some items that are not yet discounted. Although not definitive, this comparison suggests that the placement of sale signs is consistent with these predictions.

The availability of time series and cross-sectional data allows us to control for item and catalog differences. We control for item differences by comparing demand in the first and second issues with demand in the third issue. We then investigate the variance in relative demand between issues explained by sale signs. In particular, we propose the following demand function for each item in each issue:

$$Demand_{ij} = \alpha_j \gamma_i ItemPrice_{ij}^{\eta_1} OtherPrices_{ij}^{\eta_2} e^{\boldsymbol{\beta} \mathbf{X}_{ij}}. \quad (1)$$

The variable $Demand_{ij}$ represents the units sold for Item i in Issue j. The price variables $ItemPrice_{ij}$ and $OtherPrices_{ij}$ denote the price for Item i in Issue j and the average price of other items on the same page as Item i in Issue j (respectively). For sale items we used the price paid by customers rather than the discounted price. The vector $\mathbf{X}_{ij}$ measures the various marketing variables (other than price), the parameter α_j captures issue-level effects (including distribution) and the parameter γ_i captures item-level effects.

To control for the item effects (γ_i), demand in the first and second issue ($j = 1, 2$) is divided by demand in the third issue ($j = 3$). We then take logs of both sides to facilitate estimation:

$$\ln\left(\frac{Demand_{ij}}{Demand_{i3}}\right) = \hat{\alpha}_j + \eta_1 \ln\left(\frac{ItemPrice_{ij}}{ItemPrice_{i3}}\right) + \eta_2 \ln\left(\frac{OtherPrices_{ij}}{OtherPrices_{i3}}\right) + \boldsymbol{\beta}(\mathbf{X}_{ij} - \mathbf{X}_{i3}) \quad j = 1 \text{ or } 2. \quad (2)$$

In this specification the price coefficients are measures of price elasticity. Using the ratio of demand in Issues 1 and 2 relative to demand in Issue 3 controls for item-level differences (γ_i disappears). Seasonal and distribution differences between the three issues are retained in the intercepts (note that there are separate intercepts for each issue). We conclude that comparing variance in relative demand between issues provides an effective control for both issue- and item-level differences. We caution that the data do not allow us to control for possible interactions between the items and the issues. To do so would require a separate issue term for each item, which would consume too many degrees of freedom and confound some of the variables of interest.

Equation (2) is estimated using demand for 155 items that appeared in the main sections of all three issues. We omitted 7 items that appeared alone as the only item on a page because it was not possible to calculate a value for $OtherPrices_{ij}$ for these items (this omission had little effect on the coefficients of interest). Treating demand from Issues 1 and 2 separately (as a ratio of demand in Issue 3), this yielded 310 observations. We estimated coefficients and standard errors using generalized least squares (GLS) to control for the panel nature of the data. To further investigate the influence of serial correlation we also estimated separate models using relative demand in the first issue ($j = 1$) and relative demand in the second issue ($j = 2$). This approach discards information but under reasonable assumptions ensures that serial correlation cannot affect the results. Reassuringly, Hypothesis 1 is supported in these separate analyses, as the results are almost identical to those reported for the pooled analysis.

Hypothesis Testing

Recall that the objective of this analysis is to test the prediction that sale signs are less effective at increasing demand when more products have them (Hypothesis 1). We created several variables to identify this moderating effect (if any), beginning with a dummy variable identifying items with sale signs ($Sale_{ij}$). Second, we calculated the percentage of items on each page that had sale signs ($Total_{ij}$) and interacted this term with the $Sale_i$ variable. If placing a sale sign on a product increases demand for that product, we should observe a positive association between $Sale_{ij}$ and demand. If sale signs are less effective when more items have them, there should be a negative association between demand and the interaction term $Sale_{ij} * Total_{ij}$. We interpret the sign of this relationship as a test of Hypothesis 1. As we will discuss, specifying the interaction between $Sale_{ij}$ and $Total_{ij}$ in this way is statistically equivalent to including a main effect for $Total_{ij}$. However, the $(1 - Sale_{ij}) * Total_{ij}$ specification simplifies our testing of Hypothesis 1.

The alternative explanation that moderation of the sale sign effect is due to pure substitution suggests that using a sale sign may lead to substitution of demand from other products even if there is no change in the credibility of the sale signs. To explicitly control for this possibility, we identified substitute and complementary relationships between items. In particular, the variable $Substitutes_{ij}$ describes the number of other items on the same page as Item i in Issue j that are a substitute for Item i, where we defined a substitute as any item that cannot be worn at the same time as Item i. For example, two pairs of shoes would be described as substitutes because they cannot be worn together. The variable $Complements_{ij}$ describes the number of other items on the same page as Item i in Issue j that is a complement for Item i. Items are presented in the catalog by photographing them on models. We defined two items as complements if they were pictured in the same outfit worn by a model. We anticipated that demand for an item would be lower if there were more substitutes on the same page and would be higher if there were more complements. We further anticipated that these respective effects would be accentuated if a substitute or complement item had a sale sign. Therefore we constructed additional variables describing the number of complements and substitutes on the same page that had sale signs. At the suggestion of a reviewer we also investigated interacting the substitute and complement variables with the prices of other items on the page ($OtherPrices_{ij}$). These interactions explained little variance in the dependent variable and did not affect the other coefficients, and so we exclude them from the discussion. The expression $\beta\mathbf{X}_{ij}$ included the following variables and coefficients:

$$\begin{aligned}\boldsymbol{\beta}\mathbf{X}_{ij} = {} & \beta_1 Sale_{ij} + \beta_2 Sale_{ij}{*}Total_{ij} \\ & + \beta_3(1 - Sale_{ij}){*}Total_{ij} + \beta_4 Substitutes_{ij} \\ & + \beta_5 Substitutes\ on\ Sale_{ij} + \beta_6 Complements_{ij} \\ & + \beta_7 Complements\ on\ Sale_{ij} + \epsilon. \end{aligned} \tag{3}$$

Replacing $(1 - Sale_{ij}){*}Total_{ij}$ with $Total_{ij}$ yields an equivalent specification, although testing Hypothesis 1 under this alternative specification would require interpreting both β_2 and β_3. This can be seen by rearranging $\beta_2 Sale_{ij}{*}Total_{ij} + \beta_3\,(1 - Sale_{ij}){*}Total_{ij}$ as $(\beta_2 - \beta_3)\ Sale_{ij}{*}Total_{ij} + \beta_3\ Total_{ij}$. We also note that the multiplicative specification in Equation (1) anticipates an interaction between a price reduction and the presence of a sale sign. In a related study, Narasimhan et al. (1996) compare promotion elasticities across a large range of product categories. Their findings suggest that sale signs are dramatically more effective when accompanied by a price reduction, and this specification is consistent with this finding. The variables are defined as follows:

$Sale_{ij}$	1 if Item i in Issue j had a sale claim and zero otherwise;
$Total_{ij}$	Percentage of items on the same page as Item i in Issue j that have sale signs;
$Substitutes_{ij}$	The number of items on the same page as Item i in Issue j that were substitutes for that item;
Substitutes on $Sale_{ij}$	The number of items on the same page as Item i in Issue j that were substitutes for that item and had sale signs;
$Complements_{ij}$	The number of items on the same page as Item i in Issue j that were complements for that item;
Complements on $Sale_{ij}$	The number of items on the same page as Item i in Issue j that were complements for that item and had sale signs;
ϵ	Stochastic error.

Summary statistics are reported in the Appendix with a pairwise correlation table. We also investigated whether to include the claimed "regular" price to account for any reference price effect. However, this measure was heavily confounded with the price variable (the absolute correlation exceeded 0.98), and its inclusion did not influence the coefficients of interest.

Adding a sale sign to an item may affect demand for other items through both the credibility of their sale signs and any substitute or complementary relationship between the items. We summarize these effects as follows:

1. Demand for other items that already have sale signs. Under the hypothesis that sale signs are less effective when used on more items, demand for these items will decrease, reflected in a negative value for β_2. We interpret the sign of β_2 as a test of Hypothesis 1. The decrease in demand should be larger if the item receiving the additional sale sign is a substitute, suggesting that β_5 will be negative. If the item receiving the additional sale sign is a complement, then the decrease in demand should be mitigated, reflected in a positive value for β_7.

2. Demand for other items that do not have sale signs. The hypothesis that sale signs are less effective when used on more items does not speak to demand for items that do not have sale signs. We do not expect any change in the demand for these items other than through a substitute or complementary relationship with the item receiving the sale sign (reflected in β_6 and β_8), and so do not anticipate that β_3 will be significantly different from zero.

Results

We report the coefficients in Table 2 where we include three models, varying only in their inclusion of the substitute and complement controls. The significant negative coefficients estimated for $Sale_{ij}{*}Total_{ij}$ offer strong support for the prediction that sale signs are less effective when used on more products (Hypothesis 1). The $Sale_{ij}$ coefficients are positive, indicating that placing a sale sign on a product increases demand for that product, however, the negative $Sale_{ij}{*}Total_{ij}$ coefficients indicate that this effect is smaller when more of the products already have sale signs. To illustrate, we used the coefficients for Model 1 to calculate the change in demand for an item when placing a sale sign on it. If there are five items on the page, relative demand for the item receiving the sale sign increases on average by 17% if no other items

Table 2 Analysis of Historical Catalog Demand

Variables	Model 1	Model 2	Model 3
$Sale_{ij}$	0.274	0.178	0.236
	(0.138)	(0.136)	(0.144)
$Sale_{ij} * Total_{ij}$	−0.578	−0.702	−0.783
	(0.194)	(0.199)	(0.218)
$(1 - Sale_{ij}) * Total_{ij}$	0.079	0.114	0.058
	(0.110)	(0.110)	(0.157)
Log Item Price$_{ij}$	−3.235	−3.863	−3.759
	(0.463)	(0.472)	(0.481)
Log Other Prices$_{ij}$	0.735	0.620	0.618
	(0.246)	(0.242)	(0.241)
Issue 1	0.169	0.149	0.152
	(0.035)	(0.035)	(0.035)
Issue 2	0.001	−0.019	−0.016
	(0.044)	(0.043)	(0.043)
Substitutes$_{ij}$		−0.200	−0.202
		(0.046)	(0.047)
Substitutes on Sale$_{ij}$			0.018
			(0.054)
Complements$_{ij}$		0.076	−0.245
		(0.175)	(0.335)
Complements on Sale$_{ij}$			0.256
			(0.222)
Adjusted R^2	0.435	0.469	0.471
Sample size	310	310	310

Notes: Table 2 presents GLS estimates of Equation (2). Numbers in parentheses are robust standard errors (White 1980).

have sale signs and by 3% if just one other item has a sale sign. However, it does not increase at all if there are already sale signs on more than one of the other items on the page. For the sake of this illustration we assumed that the item that receives the additional sale sign is neither a complement nor a substitute of the other items on the page.

The coefficients estimated for $(1 - Sale_{ij})*Total_{ij}$ are approximately zero in all three models. The $Substitutes_{ij}$ coefficients are both significant and have their predicted signs, while the other complement and substitute variables are not significant. The $ItemPrice_{ij}$ coefficient is significant and a typical magnitude for estimates of price elasticity (Tellis 1988). The positive sign of the $OtherPrices_{ij}$ coefficient suggests that this variable can be interpreted as an additional control for substitute relationships between the items. Alternatively, high prices of other items on the page may serve as a favorable reference price, increasing demand for the focal item (Thaler 1985).

Recall that the concavity prediction does not define the domain over which concavity will be observed. The number of items with sale signs may be measured by the number on a page, the number on a facing two-page spread, or even the number in the entire catalog. In this analysis we considered the number of sale signs on the same page. As an alternative, we also considered the number of sale signs within the same two-page spread by re-estimating the model using a different version of the $Total_{ij}$ variable. We did not consider the number of sale signs in the entire issue, as any issue-level effects are confounded with other issue differences (such as distribution differences). In the spread analysis, $Total_{ij}$ measures the proportion of items that have sale signs on the same two-page spread as Item i in Issue j. The coefficients revealed similar results, although the sample size is slightly lower because we omit items on pages that shared spreads with clearance pages or the order form.

We also considered the number rather than the percentage of sale items on the page. However, the model explained less variance in relative demand, at least in part because the number of items with sale signs measure was correlated with the total number of items on the page. This introduced collinearity with the substitutes and complements measures, and resulted in a potential confound if more items on a page led to lower demand for each item.

3. Field Data

In the catalogs used to estimate the coefficients in Table 2, the assignment of sale signs and prices across items is not random. This raises an important limitation. Discussions with the catalog managers (and inspection of the data) suggest that sale signs are more likely to be placed on items that are finishing their seasons and for which there is a surplus of inventory remaining. We may be misattributing the sale sign effect if these factors contribute to the result. Problems of endogeneity and selection bias affect many studies of this type and are difficult to resolve without sufficient instruments (Chintagunta et al. 1998, Besanko et al. 1997, Villas-Boas and Winer 1996). The field study reported in this section enables us to address this limitation by replicating the findings in the historical data when simultaneously varying the use of sale signs on the same items to different randomized experimental groups.

Two experimental versions of a catalog were mailed to two separate samples of 67,000 U.S. households. The samples included both previous customers of the catalog and prospective customers identified from a purchased list. Assignment to the different conditions was approximately random, with zip codes randomly distributed across the conditions and all customers in a specific zip code receiving the same version. The catalog was mailed at the start of a season and contained a preview of that season's items, all of which are exclusive to the catalog (although competing retailers sell similar items). Recall that the credibility argument requires that customers do not know relative price levels. Using the first issue of the season in a catalog that sells exclusive items is consistent with this assumption. The items were either new to the catalog or had not appeared for almost a year (we do not have detailed item histories).

The company generously agreed to vary the use of sale claims on five items from a single product category. We cannot identify the product category (we will call it ''dresses''), but we can say that annual sales for this product category all occur in a single season. The test was conducted on four pages, positioned close to the middle of the catalog, that contained a total of eight dresses and a single accessory. In the control condition, there were ''sale signs'' on just three of the eight dresses and in the treatment condition there were sale signs on all eight dresses. *Prices for all eight dresses were held constant across both experimental conditions.* The price of the accessory was also the same in both conditions, and it did not have a sale sign in either condition. For ease of exposition we label the five dresses that had a sale sign in the treatment condition but not in the control condition as ''test dresses'' and the three dresses that had sale signs in both conditions as ''sale dresses.'' The allocation of the dresses between these two categories was determined by the catalog and was apparently arbitrary.

Sale signs were implemented by modifying the price descriptions. Prices of items with sale signs were described as follows: ''Pre-Season SALE $_.'' Prices of items without sale signs were presented with no elaboration: ''$_.'' Given the timing of the catalog at the start of the season, the ''Pre-Season SALE'' cue was an unambiguous claim that the product was new and had not previously been discounted. The price descriptions were printed in approximately 8-point type just below each item description. Because of the variety of items and stimuli in the catalog, these changes were relatively unobtrusive, minimizing the risk of demand artifacts (Shimp et al. 1991, Sawyer 1975). In total, the catalog contained approximately 300 items and more than 80 pages. Note that the three sale dresses had preseason sale claims in both conditions and included the regular price as a reference. In the treatment condition the test dresses had a preseason sale claim, but the regular price was not provided.

Results

Because the assignment of customers to the different conditions was not completely random, we compared aggregate demand for items elsewhere in the issue that did not contain experimental manipulations. The differences were negligible, and controlling for them did not affect the findings. We disguise demand by indexing it to 100 in the control condition, where demand represents the aggregate number of units sold. The results, which are presented in Table 3, replicate the findings in §2 and provide further evidence that sale signs are less effective when more products have them (Hypothesis 1). Demand for the three sale dresses was significantly ($p < 0.05$) higher in the control condition, where they were the only items with sale signs, than in the treatment condition where the test dresses also had sale signs. As expected, placing a sale sign on the test dresses in the treatment condition significantly increased demand for these dresses compared to the control condition. There was almost no variance in orders for the accessory.

The findings in Table 3 are consistent with the find-

Table 3 Results of the Field Test

Item	Treatment	Control	Prediction	Treatment	Control
Test dresses	Sale sign	No sale sign	Higher demand in the treatment condition	157*	100
Sale dresses	Sale sign	Sale sign	Lower demand in the treatment condition	38*	100
Accessory	No sale sign	No sale sign	No difference between conditions	90	100

Notes: Table 3 describes the number of units purchased in each experimental condition aggregated across the respective items. The numbers in each row are indexed to 100 in the control condition (it is not appropriate to compare between rows). For example, the catalog sold 57% more of the five test dresses in the treatment condition than in the control condition.
*Demand in the treatment condition was significantly higher (lower) than demand in the control condition ($p < 0.05$).

ing from the historical data that sale signs are more effective when fewer products have them. Recall that a major limitation of that finding was that the assignment of sale signs and prices across items in the historical data was not random, introducing problems of endogeneity and selection bias. These concerns do not apply to the field test, and so it is reassuring that the finding still holds. However, varying the content of catalogs is expensive (this study apparently cost over $20,000) and preserving the cooperation of a catalog retailer restricts discretion over the experimental design. As a result, the field study has its own limitations. First, we cannot rule out the possibility that the effects are limited to the test items or the dress category. Second, it is possible, though unlikely, that the omission of the regular price from the price descriptions for the test dresses in the treatment condition contributed to the diminished demand for the sale dresses in that condition. Finally, perhaps the most serious limitation is that the results in Table 3 are consistent with both the credibility theory and a substitution explanation. Placing sale signs on the test dresses may have reduced demand for the sale dresses purely through substitution. Therefore, our finding that demand for the sale dresses was lower when the test dresses had sale signs (treatment condition) is consistent with, but does not necessarily imply, that the sale signs on the sale dresses were less credible in this condition. The substitution explanation does not apply to the historical data analyzed in §2. In particular, the historical data analysis described the impact of varying sale signs on a much larger variety of products and product categories. Moreover, we explicitly controlled for substitute and complementary relationships between the products. We conclude that it is reassuring that the analyses do not share the same limitations yet they yield similar results.

It is not possible to calculate aggregate demand for all eight dresses from Table 3 (recall that demand is standardized within each row). However, we can report that there was no significant difference in aggregate demand for the eight dresses across the two conditions. This raises the issue of total category sales. Recall that the argument predicts that the relationship between aggregate demand and the number of items with sale signs will be an inverted U (Hypothesis 2). Because we only compare two conditions, there is no prediction about whether demand for all items will be higher when eight dresses have sale signs compared to when just three dresses have sale signs. Testing the prediction regarding total category demand (Hypothesis 2) in this field test would have required multiple comparisons varying the presence of sale signs across all of the dresses in the catalog (rather than just the dresses on these four pages). Instead, the field test focuses on Hypothesis 1, and we introduce grocery store data in the next section to test the category sales prediction (Hypothesis 2). As we will discuss, these data provide a strong external control for substitution and help to evaluate whether the findings in these first two sections are generalizable to other markets and retail formats.

4. Grocery Store Data

In this section we test whether total category demand is maximized when some but not all of the items in the category have a sale sign (Hypothesis 2). The credibility argument predicts that the benefit from placing a sale sign on an additional product is eventually outweighed by the loss of credibility among the existing sale signs. This implies that adding another sale sign will eventually *decrease* total category sales when many items already have sale signs. We test this prediction using grocery store scan data describing total category demand for frozen juice, toothpaste, and canned tuna (all canned fish) at Dominick's Finer Foods grocery stores in Chicago.

The analysis allows us to discriminate between the credibility argument and a pure substitution explanation for the evidence in the previous two sections that demand for items with sale signs is smaller when more items have sale signs. Total category demand provides an effective control for substitution within the category. If adding a sale sign to an item leads solely to substitution from other products within the category we will not observe a decrease in total category demand. We caution that this does not control for substitution between categories. However, if using more sale signs within the category makes these items more attractive, we do not expect cross-category substitution to cause category contraction.

The analysis also helps to evaluate whether the results from the catalog data can be generalized to other markets. Frozen juice, tuna, and toothpaste are frequently purchased products that are homogeneous across stores. These characteristics are likely to result in customers being more informed about prices for these products than about prices for fashion clothing. However, a series of behavioral studies investigating customers' price knowledge of grocery products suggests that even in grocery stores many customers are poorly informed about prices. Most of these studies reveal that no more than half of customers questioned can recall the prices of products that they have recently purchased (Allen et al. 1976, Conover 1986, Dickson and Sawyer 1990).

There are other differences between our previous analysis and the analysis in this section that are worthy of discussion. First, prices in fashion markets vary over time due to seasonality in demand. However, the source of price variance is not important to their model, and so their findings should apply equally to markets in which prices vary due to supply shocks, such as manufacturers' trade discounts. Second, in the previous analysis we evaluated the number of items on a page or a two-page spread, while for the grocery markets we consider the number of items in the category. As we discussed, the theory is silent as to the domain over which customers judge credibility, and so this domain remains an empirical question. Third, increases in category sales of frozen juice and tuna reflect a combination of purchase acceleration and category expansion. Category expansion is less likely with toothpaste, and so increases in category demand for toothpaste are more likely to be due to purchase acceleration (note that none of these products is perishable). Because customers may purchase from a competing store in the future, both purchase acceleration and category expansion (while controlling for price changes) contribute directly to retailer profits.

The data include weekly observations of total sales in each category for 169 weeks at approximately 80 stores (the number of stores varies by category, see Appendix). A complete description of the data for each category, including the individual UPCs, is available at http://gsbwww.uchicago.edu/research/mkt/Databases/DFF/W.html. The data include variables indicating whether an item was sold on promotion that week, together with the number of units sold, the price of each item, and whether a major holiday fell during that week. The nature of the variables is described in some detail on the University of Chicago Web site. Unfortunately, the description of the promotion variable is not very complete. However, we interpret the variable in the same way as Kadiyali et al. (2000). The promotion variable does appear to describe in-store promotion rather than just advertising.

Hypothesis Testing

The objective of this analysis is to test whether category sales are maximized when some but not all products have sale signs. Because this prediction suggests a nonlinear relationship between category sales and the number of items with sale signs, we created a series of dummy variables that identify how many items in the category have sale signs. For example, the variable $Number\ 0\text{–}5_{ij}$ identifies observations (store-weeks) for which between zero and five items in the category have sale signs, while *Number 6–10*$_{ij}$ identifies observations with between six and ten sale signs in the category. Using dummy variables to specify the number of items with sale signs eliminates the need to fully specify the shape of the response function in advance. Instead, the model finds the functional form that best fits the data. Constructing dummy variables to represent intervals of five ensures that there are sufficient observations to reliably estimate coefficients for each interval. As we would expect, the findings are robust to varying the intervals over which the dummy variables are constructed.

Recall that the theory is silent as to the domain over which customers judge credibility, and does not describe whether customers are sensitive to the *number* or *percentage* of sale signs in a category (or both). The relevant domain is an empirical question, which we address by considering both domains. We constructed a second set of dummy variables to identify the proportion of items with sale signs in increments of 5%. For example, the variable *Percent 6–10*$_{ij}$ identifies observations where between 6% and 10% of the items in the category had sale signs. For completeness, we separately report the findings for both the *Percent* and *Number* variables.

Which items had sale signs is also likely to affect total category demand. We would expect that placing a sale sign on more popular brands and larger package sizes will lead to more purchase acceleration and have a bigger impact on aggregate demand than placing a sale sign on a less preferred item. To control for this variance, we calculated the average market share of items with sale signs in each week at each store (*Share of Sale Items*$_{ij}$). To lessen the possibility that aggregation will mask any price changes that coincide with the use of sale signs, we incorporate separate price vectors for each of the major brands. The price vectors were calculated as the average price for Brand *b* in each week at each store, weighted by the market share of each SKU (*Price*$_{ijb}$). An "other" price vector was calculated for the minor brands. To avoid introducing confounds with the weekly sales measure, we used the entire data period to calculate the market share measures used in the *Price*$_{ijb}$ and *Share of Sale Items*$_{ij}$ variables.

Other variables measured the number of items in the category in each store each week (*Number of Items*$_{ij}$) and weeks in which major holidays fell (*Holiday*$_i$). To calculate total category sales (*Sales*$_{ij}$) we aggregated across different package sizes using the net weight of the package. We summarize these variables as follows:

Sales$_{ij}$	Total category sales for Week *i* at Store *j*;
Price$_{ijb}$	Average category price for Brand *b* in Week *i* at Store *j*;
Share of Sale Items$_{ij}$	Average market share of items with sale signs in Week *i* at Store *j*;
Number of Items$_{ij}$	Number of unique brand-sizes in Week *i* at Store *j*;
Holiday$_i$	1 if a major holiday fell in Week *i* and zero otherwise;
Store$_j$	1 if Store *j* and zero otherwise.

Summary statistics and a correlation matrix are provided in the appendix. Although not directly observable in the category-level data, analysis of the item-level data confirms that there is a significant negative association between prices and sale signs, indicating that sale signs are informative about which prices have been discounted. Aggregating up to the brand level, we also replicated the analysis conducted on the historical catalog data (§2). This replication yielded strong evidence across all three categories that sale signs are less effective at increasing demand when they are used on more products (Hypothesis 1).

Adopting the same specification as the demand function used in our analysis of historical sales data from the women's clothing catalog, we estimated the following equation:

$$\ln Sales_{ij} = \sum_{j=1}^{J} \alpha_j Store_j + \sum_{b=1}^{B} \eta_b \ln Price_{ijb} + \boldsymbol{\beta}\mathbf{X}_{ij}. \quad (4)$$

The store coefficients account for geographical fluctuations in demand. To account for the panel nature of the data we estimated the parameters using GLS and used White's (1980) method to correct the standard errors. The $\boldsymbol{\beta}\mathbf{X}_{ij}$ expression included the following terms:

Table 4a Analysis of Grocery Store Data: Frozen Juice

	Coefficient		Coefficient
Number 6–10	0.247	Percent 6–10	0.177
	(0.011)		(0.016)
Number 11–15	0.371	Percent 11–15	0.292
	(0.012)		(0.016)
Number 16–20	0.339	Percent 16–20	0.394
	(0.012)		(0.016)
Number over 20	0.279	Percent 21–25	0.383
	(0.014)		(0.016)
		Percent 26–30	0.382
			(0.017)
		Percent 31–35	0.299
			(0.018)
		Percent over 35	0.238
			(0.020)
Other Variables			
Log Tropicana Price	−0.088		−0.079
	(0.019)		(0.019)
Log Minute Maid Price	−0.062		−0.123
	(0.027)		(0.027)
Log Store Brand Price	−0.601		−0.594
	(0.033)		(0.034)
Log Dole Price	−0.235		−0.295
	(0.030)		(0.030)
Log Welch's Price	−0.069		−0.059
	(0.053)		(0.053)
Log Other Price	−0.300		−0.267
	(0.057)		(0.057)
Share of Sale Items	10.853		10.622
	(0.401)		(0.399)
Number of Items	−0.015		−0.015
	(0.001)		(0.001)
Holiday	−0.005		0.018
	(0.008)		(0.008)
Intercept	7.608		7.489
	(0.160)		(0.161)
R^2	0.59		0.59
Sample size	12199		12199

Notes: Table 4a presents GLS estimates of Equation (5) for the frozen juice data. Store-specific coefficients are omitted. Numbers in parentheses are robust standard errors (White 1980).

Table 4b Analysis of Grocery Store Data: Toothpaste

	Coefficient		Coefficient
Number 6–10	0.040	Percent 6–10	0.032
	(0.004)		(0.004)
Number 11–15	0.044	Percent 11–15	0.049
	(0.004)		(0.004)
Number 16–20	0.069	Percent 16–20	0.045
	(0.006)		(0.005)
Number over 20	0.022	Percent 21–25	0.079
	(0.007)		(0.007)
		Percent 26–30	0.038
			(0.012)
		Percent 31–35	0.012
			(0.017)
		Percent over 35	0.001
			(0.022)
Other Variables			
Log Crest Price	−0.287		−0.294
	(0.037)		(0.037)
Log Colgate Price	−0.586		−0.596
	(0.029)		(0.029)
Log Aquafresh Price	−0.116		−0.116
	(0.025)		(0.025)
Log Close Up Price	−0.105		−0.096
	(0.022)		(0.022)
Log Other Price	0.006		0.002
	(0.015)		(0.015)
Share of Sale Items	4.218		4.031
	(0.433)		(0.439)
Number of Items	0.010		0.010
	(0.000)		(0.000)
Holiday	0.002		0.004
	(0.004)		(0.004)
Intercept	5.663		5.644
	(0.052)		(0.051)
R^2	0.75		0.78
Sample size	12404		12404

Notes: Table 4b presents GLS estimates of Equation (5) for the toothpaste data. Store-specific coefficients are omitted. Numbers in parentheses are robust standard errors (White 1980).

$$\boldsymbol{\beta}\mathbf{X}_{ij} = \sum_{g=1}^{G} \beta_g \mathit{Number}_g + \tau_1 \mathit{Share\ of\ Sale\ Items}_{ij} + \tau_2 \mathit{Number\ of\ Items}_{ij} + \tau_3 \mathit{Holiday}_i + \epsilon. \quad (5)$$

We use the subscript g to distinguish between the dummy variables identifying the number of items with sale signs. The first variable *Number 0–5*$_{ij}$ is omitted, so that the variables describe changes in category demand above or below this base level. In the percentage analysis, the *Number*$_g$ variables are replaced with *Percent*$_g$ variables (omitting *Percent 0–5*$_{ij}$).

Results

The findings for each category are reported in Tables 4a, 4b, and 4c. The coefficients are consistent with the

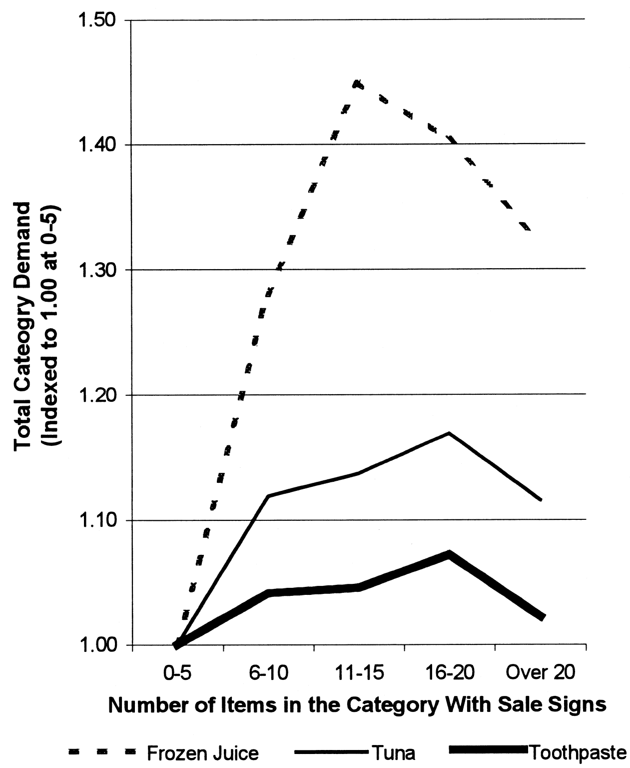
Table 4c Analysis of Grocery Store Data: Tuna

	Coefficient		Coefficient
Number 6–10	0.112	Percent 6–10	0.008
	(0.008)		(0.017)
Number 11–15	0.128	Percent 11–15	0.040
	(0.009)		(0.016)
Number 16–20	0.156	Percent 16–20	0.036
	(0.014)		(0.016)
Number over 20	0.109	Percent 21–25	0.172
	(0.023)		(0.016)
		Percent 26–30	0.154
			(0.017)
		Percent 31–35	0.128
			(0.018)
		Percent over 35	0.134
			(0.018)
Other Variables			
Log Chicken of the Sea Price	−1.397		−1.333
	(0.036)		(0.036)
Log Starkist price	−0.688		−0.676
	(0.039)		(0.039)
Log Bumble Bee Price	−0.113		−0.094
	(0.039)		(0.039)
Log Other Price	−0.194		−0.128
	(0.056)		(0.056)
Share of Sale Items	−2.049		−1.867
	(0.270)		(0.270)
Number of Items	−0.014		−0.013
	(0.001)		(0.001)
Holiday	−0.127		−0.125
	(0.008)		(0.008)
Intercept	5.316		5.525
	(0.152)		(0.154)
R^2	0.47		0.48
Sample size	13914		13914

Notes: Table 4c presents GLS estimates of Equation (5) for the canned tuna data. Store-specific coefficients are omitted. Numbers in parentheses are robust standard errors (White 1980).

prediction that category sales are highest when some, but not all, products have sale signs (Hypothesis 2). Category demand in each category peaked when approximately 25% of the items in the category had sale signs. Notably, in all three categories there is a significant reduction in aggregate demand from the respective peaks when more than 30% of items had sale signs. We depict the relationship between total category sales and the number of items with sale signs in Figure 1. For ease of comparison, we index the curves to one when fewer than six items have sale signs. The curves are calculated using the coefficients reported in Tables 4a, 4b, and 4c, together with the averages of the Log $Price_{ijb}$, $Number\ of\ Items_{ij}$, $Share\ of\ Sale\ Items_{ij}$, and $Holiday_i$ variables (see the appendix).

Figure 1 Total Category Demand and the Number of Items with Sale Signs



The curves suggest that sale signs have a larger impact on the aggregate demand for frozen juice and tuna than on the demand for toothpaste. This is consistent with our earlier observation that category expansion is less likely with toothpaste. We also note that, on average, stores used fewer sale signs than the level at which category demand peaks. One explanation for this discrepancy is that the number of sale signs used in these categories affects the credibility of other sale signs in the store. This will tend to result in stores using fewer sale signs in a category than required to maximize demand within that category. Consistent with this interpretation, Anderson and Simester (1998) describe a department store that has a policy limiting the number of sale signs in any one department. One would normally expect that individual department managers would be best positioned to choose how many sale signs to use in their department. A corporate policy limiting the discretion of department managers suggests that there are externalities that individual department managers overlook. Department managers may optimize the use of sale signs for their own departments, but fail to internalize the impact of additional sale signs on the credibility of sale signs in other departments. This leads to an outcome that is similar to the grocery store findings, with the store placing sale signs on fewer products within each department than is required to maximize demand for that department.

We caution that the findings in Figure 1 focus on the role of sale signs, not prices. The model controls separately for price changes using a series of brand price vectors, and the coefficients estimated for these price vectors are negative. We cannot conclude that adding sales signs, accompanied by price reductions, will decrease category sales. Total category sales may continue to expand due to the simultaneous reduction in price, which always increases sales, and the addition of a sale sign, which may or may not increase sales.

In other findings of interest, the $Holiday_{ij}$ variable is negative (where significant), indicating that demand is lower during holiday periods for frozen juice and tuna, although the demand for toothpaste is apparently unaffected. We also considered a variable identifying weeks prior to a major holiday; however, the variable explained little variance and did not affect the other coefficients of interest. The $Number\ of\ Items_{ij}$ and $Share\ of\ Sale\ Items_{ij}$ are significant in all three models, although their magnitudes and signs vary, indicating that these effects are specific to each category. Reassuringly, the finding that category sales are highest when some but not all products have sale signs is just as strong if these variables are omitted.

The reduction in category sales when sale signs are more prevalent cannot be explained by substitution within the category, but is consistent with the prediction that sale signs are less credible when used on more products. The results indicate that the benefit from placing a sale sign on an additional product is eventually outweighed by the loss of credibility among the existing sale signs. Like the analysis of the historical data from the catalog, the major limitation of this analysis is the endogeneity in the number of items that have sale signs. If the events that prompt variance in the number of items with sale signs also lead to variance in actual demand, we may be incorrectly attributing the change in demand. Although it is not possible to rule out this limitation, it is difficult to identify events that would explain both an initial increase and a subsequent decrease in category demand as the number of sale signs increases.

5. Customer Beliefs

In this section we use survey measures to evaluate Hypothesis 3, which predicts both that placing a sale sign on a product reduces the perceived likelihood that the product will be available at a lower price in the next period, and that this effect is smaller when more items have signs. The analysis helps us to discriminate between the credibility explanation and an attention effect. Distinguishing between the two explanations is a difficult task. They both imply that sale signs deliver less information when there are too many signs. The credibility argument predicts that this occurs because the signs are noticed but not believed. The attention explanation predicts that the signs are less likely to be noticed when attention is diluted by a large number of sale signs. We attempt to control for the attention effect by focusing subjects' attention on a set of focal items. Under these conditions, it is unlikely that subjects overlook sale signs on these items. More important, the likelihood of overlooking sale signs on the focal items is unlikely to depend on how many other items also have sale signs.

Respondents were presented with a two-page questionnaire, with a copy of either a treatment or control version of a survey stimulus designed to portray pages from an actual catalog. The surveys were completed by a convenience sample of 79 undergraduate student respondents recruited from the campus of a large university. Potential respondents were approached by a research assistant in the university library and asked whether they would be willing to complete a short survey. The research assistant, who was not informed about the goal of the study, reported that nearly all of the potential respondents who were approached agreed to participate. The respondents were paid $2 for their participation.

The survey stimulus contained four pages with four clothing products on each page, all selected from the pages of a clothing catalog whose target market includes students (the respondent population). Color images of the products were taken directly from the company's Web site. The 16 products were chosen from different pages of the catalog and from different product categories to minimize the likelihood that they were either substitutes or complements. The four pages were then printed in color and arranged in a six-page booklet, with the first and last pages left blank. The prices for all 16 items were the same in both versions of the catalog. The only difference between the versions was the number of sale signs on each page. In particular, the price descriptions in the treatment version for 3 of the 4 items on each page indicated that the prices were discounted, as follows: "SALE $X." Prices of the other item were presented simply as: "$X." In the control version, the sale cue appeared on just 1 item on each page (these items had sale cues in both versions). Given this design, we can categorize the 16 items into three groups: 8 items that had sale cues only in the treatment version (test group), 4 items that had sale cues in both versions (sale group), and 4 items that did not have sale cues in either version (nonsale group).

Instructions at the start of the questionnaire advised the respondents that the stimulus was a replica of a current catalog and asked them to briefly browse through the catalog. For three items the subjects were then asked both the likelihood that an item would be available in the future and the expected future price of the item if it was available. Asking questions about a small number of items ensured that these items received attention and made it unlikely that subjects would overlook whether they had sale signs. The possibility that subjects did not notice sale signs on other items reduces the difference between the conditions, making it less likely that we will observe significant differences in responses between conditions (making the test more conservative). The actual wording of the availability question was as follows: "If the next issue of this catalog is mailed out two months later, how likely is the ⟨item⟩ to be available in the next issue."

Responses were collected on an 11-point semantic differential scale anchored at 0 by *Definitely not available* and at 10 by *Definitely available.* The future price question instructed respondents to assume that the item was available in the next issue and asked them: "What is your best guess of the price in the next issue of the catalog?"

Respondents were asked to specify an actual price. No time limit was placed on the subjects, and they were able to freely examine the catalog while completing the survey. Extensive pretesting confirmed that the questions were unambiguous and subjects reported no difficulty in responding to them. The three items were all located on different pages of the stimulus and were randomly selected, with one item from each of the three groups. For ease of exposition we label them test item, sale item, and nonsale item, corresponding to the group from which they were drawn.

The theory is not explicit regarding the time frame of customers' expectations. In Anderson and Simester's model (1998), products have two period seasons, and customers are initially unsure whether a product is in the first or second period of its season. In this study we operationalize the time between periods as the time between issues of a catalog. As we discussed in the Introduction, these difficulties operationalizing the model's parameters raise the risk that the predictions will be rejected even if they hold in other settings.

Hypothesis Testing

Recall that the objective of this analysis is to test the prediction that placing a sale sign on a product reduces the perceived likelihood that the product will be available at a lower price in the future, and that this effect is smaller when more items have signs. The measurement of each effect mimics the design used in the field study (§3). First, to evaluate whether placing a sale sign on a product reduces the perceived

Table 5 Difference Between Treatment and Control

Item	Treatment	Control	Prediction	Likely to Be Available[a]	Future Price[b]
Test item	Sale sign	No sale sign	Less likely to be available and higher future price (if available) in treatment condition	−0.97*	13%**
Sale item	Sale sign	Sale sign	More likely to be available and lower future price (if available) in treatment condition	2.34**	−6%*
Nonsale item	No sale sign	No sale sign	No difference between conditions	−0.46	1%

[a]The treament group average minus the control group average. A positive (negative) value indicates a higher (lower) likelihood of future availability in the treament condition than in the control condition.
[b]The treament group average minus the control group as a percentage of the control group average. A positive (negative) value indicates a higher (lower) future price in the treament condition than in the control condition.
**Average significantly larger (smaller) in the treatment group than in the control ($p < .01$).
*Average significantly larger (smaller) in the treatment group than in the control ($p < .05$).

likelihood that the product will be available at a lower price in the future, we compare the response in each condition for the test item. This item has a sale sign in the treatment condition but no sale sign in the control, and serves an analogous role to the test dresses in the field test.

Second, to evaluate whether the number of items with sale signs moderates the effect, we compare the response in each condition for the sale item, which has a sale sign in both conditions. This item serves an analogous role to the sale dresses in the field test. Recall that in the field test we compared whether the demand for items with sale signs is smaller when more items have them. In this study we compare whether subjects believe that it is more likely that an item with a sale sign will be available at a lower price in the future when more items have sale signs. In particular, there are 12 items with sale signs in the treatment condition compared with 4 items in the control condition. If sale signs are less effective in the treatment condition, subjects in that condition should believe that the sale item is more likely to be available in the future and, if available, more likely to have a lower price than should subjects in the control condition. Formally, let expectations about the future availability of this item be $E - \Delta$, where E are subjects' expectations without sale signs and Δ represents the sale sign effect. The prediction is that $E - \Delta_T > E - \Delta_C$, which allows us to evaluate whether $\Delta_T < \Delta_C$, where the subscripts T and C denote treatment and control, respectively.

Finally, as with the accessory in the field test, the theory does not predict any difference between conditions in the response to items that do not have sale signs in either condition. Therefore, for the nonsale item, which does not have a sale sign in either condition, we expect to observe the same average response to the availability and future price in each condition.

Results

A total of 39 respondents saw the control version of the catalog (sale signs on 4 items) and 40 respondents saw the treatment version (sale signs on 12 items). The results, which are summarized in Table 5, offer support for Hypothesis 3. The presence of a sale sign reduced expectations that a product would be available at a lower price in the future. Respondents in the treatment condition all saw a sale sign on the test item and indicated on average that it was both less likely to be available in the future and, if available, expected it to have a higher price compared to respondents in the control condition. Also consistent with Hypothesis 3, the effectiveness of sale signs appears to diminish when more products have sale signs. All respondents saw sale signs on the sale item, but those in the treatment condition saw sale signs on 75% of the items, compared with respondents in the control condition who saw sale signs on only 25% of the items. The findings for the sale item indicate that, on average, subjects believe it is more likely that an item with a sale sign will be available at a lower price in the future if more items have sale signs.

We did not observe any significant differences between conditions in the average response to either question for the nonsale item. This null result is reassuring, and suggests that the respondents in the two groups were equivalent and that their price and availability expectations for this item were not affected by the experimental manipulations.

We caution that we did not collect process measures to evaluate how many customers were aware of the presence or absence of sale signs on the three focal items. For this reason, we are not able to conclusively rule out the possibility that the support for Hypothesis 3 is due to more customers noticing the sale signs on the sale item in the control condition (where fewer items had sale signs). Although we believe that this is unlikely, further research directed at this issue would complement these results.

6. Other Implications

The findings have broader implications for the estimation of choice models, the selection of retail promotion strategies, and the testing of signaling models. In this section we discuss these and other implications.

We earlier acknowledged that the discrete choice literature offers persuasive evidence that placing a sale sign or other promotional cue on an item can increase demand for that item (Guadagni and Little 1983, Grover and Srinivasan 1989). The findings suggest that the specification of discrete choice models can be improved by recognizing that the effectiveness of sale signs is moderated by the number of items that have them. To illustrate the importance of this interaction we estimated sales at the brand level using the tuna, toothpaste, and frozen juice data described in §4. Consistent with our previous results, we found strong support for Hypothesis 1. In each category the total number of sale signs in the category significantly moderates the marginal value of adding a sale sign to an item.

The data in this paper focus on sale signs placed on individual items, yet stores also engage in storewide sales events, such as grand opening, Anniversary, and one-day sales. Salop (1977) and others investigate the use of periodic sales by monopolists. This work suggests that storewide events such as one-day sales enable retailers to price-discriminate between customers who are time sensitive and other customers who are time insensitive. The findings in this paper complement their work by predicting that these events will be more effective when they occur less frequently. Customers may question the credibility of frequent storewide events (e.g., anniversary sales that occur more than once a year).

The findings also relate to Thaler's (1985) model of consumer choice. Thaler introduces the notion of transaction utility, reflecting the utility derived from participating in a favorable transaction. More formally, he defines transaction utility as a function of the difference between the price paid and some reference price. The lower the price paid relative to the reference price, the larger the transaction utility. The findings reported here may be interpreted as evidence that sale signs provide a credible signal that raises customers' references prices, which in turn raises their transaction utility. We note that Thaler's model does not predict which cues affect a customer's reference price or explain why such cues are credible. The two explanations are better thought of as complementary rather than competing.

While the present paper focuses on sale signs, our empirical results may apply more broadly to other types of marketing-mix elements, such as price endings, promotional displays, and weekly store feature advertisements. When marketers learn that these cues are effective, there is a natural temptation to increase usage, yet the effectiveness of the cues may be moderated by use. This suggests a similar trade-off to that documented for sale signs: Increased usage of the marketing-mix elements will benefit those items that receive additional marketing support, but may decrease demand for those items already receiving marketing support. For example, Berman and Evans (1992) and Mason and Mayer (1990) suggest that the demand premium associated with 9-digit price endings, such as \$9 or 99¢, may result from customers interpreting these price endings as an indication that a price is low relative to other market prices (see Stiving and Winer 1997 for a review of the price endings literature). Under this interpretation, 9-digit price endings serve an analogous role to sale signs. The findings in this paper suggest that the effectiveness of 9-digit price endings will diminish when they are used on more items.

Behavioral studies have investigated customers' price knowledge of grocery products. As we discussed, most of these studies reveal that no more than half of the customers questioned can recall the prices of recently purchased items (Allen et al. 1976, Conover 1986, Dickson and Sawyer 1990). The findings in this paper may help to explain these findings. If sale signs are informative, customers need not remember market price levels to evaluate whether to visit a competing store or delay in the hope of a future discount. Instead, customers can rely on sale signs to help guide their decisions. We conclude that customers who are poorly informed about market prices may represent an implication of the Anderson and Simester (1998) model, rather than an assumption.

7. Conclusions

We have presented data from several sources, including historical data from a women's clothing catalog, a field study in that catalog, grocery store sales data for frozen juice, toothpaste, and tuna; and survey responses to catalog stimuli. Analysis of these data yields three conclusions. First, sale signs are less effective at increasing demand when more items have them. Second, total category sales are maximized when some but not all products have sale signs. Third, placing a sale sign on a product reduces the perceived likelihood that the product will be available at a lower price in the future, but the effect is smaller when more products have sale signs. By focusing on settings that allows us to rule out alternative explanations, the findings suggest that the moderation of the sale sign effect is in part due to reduced credibility when they are used on more products. We caution that there may exist other circumstances in which these alternative explanations lead to similar findings. The consistency of these findings is particularly reassuring given that limitations in the data do not extend across all of the data sources.

The results also offer empirical insights that have more general application than testing the credibility theory. Promotional cues are important marketing-mix variables; adding a sale sign can have a demand effect exceeding 50% without changing the price. The evidence in this paper helps explain why they are so effective, and cautions against overuse.

Acknowledgments
The paper has benefited from comments by Dan Ariely, Pradeep Chintagunta, John Hauser, Nanda Kumar, Nader Tavassoli, Birger Wernerfelt, reviewers and editors of this journal, together with seminar participants at Chicago, Duke, MIT, Notre Dame, Rochester, Wharton, and the University of New South Wales. We also wish to acknowledge generous contributions of the mail order catalog company that provided data for this study and the Marketing Group at the GSB, University of Chicago, for the grocery store data.

Appendix

Historical Catalog Data Summary Statistics

	1st and 2nd Issues		3rd Issue	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*
Item price	63.97	30.31	62.09	30.47
Items with sale signs (%)	9.27		25.6	
Number of items of each page	4.79	2.87	4.88	2.47
Substitutes	1.48	1.42	1.61	1.40
Substitutes with sale signs	0.08	0.28	0.45	0.58
Complements	0.20	0.48	0.22	0.49
Complements with sale signs	0.03	0.17	0.06	0.27
Sample size	155		155	

Pairwise Pearson Correlation Coefficient of Historical Catalog Data Variables ($N = 162$)

	Log Quantity	Log Price	Sale	Sale * Total	(1 − Sale) * Total	Substitutes	Substitutes on Sale	Complements
Log Price	−0.578**							
Sale	0.518**	−0.844**						
Sale ∗ Total	0.344**	−0.658**	0.818**					
(1 − Sale) ∗ Total	−0.139	0.248**	−0.300**	−0.329**				
Substitutes	−0.176*	−0.081	−0.067	−0.112	0.199*			
Substitutes on Sale	−0.096	0.072	−0.120	−0.056	0.686**	0.290**		
Complements	0.022	−0.094	0.177*	0.288**	0.116	−0.120	0.073	
Complements on Sale	0.021	−0.077	0.090	0.258**	0.164*	−0.093	0.113	0.870**

*$p < 0.05$. **$p < 0.01$.

Grocery Store Data Summary Statistics

	Frozen Juice		Toothpaste		Tuna	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*
Log sales	9.99	0.46	7.84	0.35	8.96	0.50
Share of sale items	1.59	0.802	0.59	0.39	2.25	1.19
Number of items	69.09	3.69	90.13	17.49	44.93	4.61
Number of items with sale signs	13.21	6.43	9.13	6.96	9.41	5.09
Holiday	0.17	0.38	0.17	0.38	0.17	0.38
Number of stores	78		77		86	
Samples size	12,199		12,404		13,911	

Note: Missing observations reflect the absence of stores from the sample.

Pairwise Pearson Correlation Coefficients of Frozen Juice Variables

	Log Sales	Share of Sale Items	Number of Items	Number of Items with sale signs	Percentage of Items with sale signs	Holiday
Share of sale items	0.44*					
Number of items	0.217*	0.022*				
Number of items with sale signs	0.058*	−0.352*	0.017			
Percent of items with sale signs	0.032*	−0.354*	−0.095*	0.992*		
Holiday	0.008	−0.018	−0.026*	0.229*	0.229*	
Log Tropicana Price	−0.020*	−0.151*	−0.128*	0.006	0.021	−0.092*
Log Minute Maid Price	0.006	0.017	−0.179*	−0.200*	−0.182*	−0.071*
Log Store Brand Price	−0.068*	−0.098*	−0.211*	0.066*	0.091*	−0.028*
Log Dole Price	−0.030*	0.111*	−0.019	−0.258*	−0.256*	0.009
Log Welch's Price	0.041*	0.150*	0.043*	−0.198*	−0.205*	−0.039*
Log Other Price	0.075*	0.058*	0.213*	−0.031*	−0.055*	−0.034*

*$p < 0.01$.

Pairwise Pearson Correlation Coefficients of Toothpaste Variables

	Log Sales	Share of Sale items	Number of Items	Number of Items with sale signs	Percentage of Items with sale signs	Holiday
Share of sale items	−0.051*					
Number of items	0.564*	−0.178*				
Number of items with sale signs	0.187*	0.151*	0.211*			
Percent of items with sale signs	0.052*	0.204*	−0.041*	0.950*		
Holiday	−0.018*	0.040*	−0.021	−0.075*	−0.076*	
Log Crest Price	−0.119*	0.007	0.134*	−0.031*	−0.074*	−0.001
Log Colgate Price	−0.022*	−0.075*	0.277*	−0.052*	−0.134*	0.025*
Log Aquafresh Price	−0.108*	0.135*	0.122*	0.029	−0.011	0.011*
Log Close Up Price	−0.122*	0.091*	−0.024*	−0.016	−0.022	0.017*
Log Other Price	0.064*	−0.191*	0.066*	−0.129*	−0.160*	−0.046*

*$p < 0.01$.

Pairwise Pearson Correlation Coefficients of Tuna Variables

	Log Sales	Share of Sale Items	Number of Items	Number of Items with Sale Signs	Percentage of Items with Sale Signs	Holiday
Share of sale items	−0.051*					
Number of items	0.028*	0.026*				
Number of items with sale signs	0.204*	0.025*	0.208*			
Percent of items with sale signs	0.209*	0.028*	0.031*	0.979*		
Holiday	−0.102*	−0.053*	−0.052*	0.002	0.018	
Log Chicken of the Sea Price	−0.218*	−0.014	−0.053*	−0.209*	−0.212*	0.078*
Log Starkist Price	−0.070*	−0.093*	0.026	−0.172*	−0.189*	−0.043*
Log Bumble Bee Price	0.032*	−0.071*	0.077*	−0.066*	−0.087*	0.008
Log Other Price	−0.007	0.010	0.170*	−0.070*	−0.108*	−0.017

*$p < 0.01$.

References

Allen, J. W., G. D. Harrell, M. D. Hutt. 1976. *Price Awareness Study.* The Food Marketing Institute, Washington D.C.

Anderson, Eric T., Duncan I. Simester. 1998. The role of sale signs. *Marketing Sci.* **17**(2) 139–155.

Berman, Barry, Joel R. Evans. 1992. *Retail Management: A Strategic Approach.* Macmillan, New York.

Besanko, David, Sachin Gupta, Dipak Jain. 1997. Logit demand estimation under competitive pricing behavior: An equilibrium framework. Working paper, Northwestern University, Evanston, IL.

Chintagunta, Pradeep K., Vrinda Kadiyali, Naufel J. Vilcassim. 1998. Competitive interactions in price and advertising: An analysis of the 'share of voice to share of market' rule. Working paper, University of Chicago, Chicago, IL.

Conover, Jerry. 1986. The accuracy of price knowledge: Issues in research methodology. Richard Lutz, ed. *Advances in Consumer Research,* vol. 13. Association for Consumer Research, Ann Arbor, MI, 589–593.

Dickson, Peter, Alan Sawyer. 1990. The price knowledge and search of supermarket shoppers. *J. Marketing* **54** (July) 42–53.

Grover, Rajiv, V. Srinivasan. 1989. Evaluating the effects of retail promotions in segmented markets. Working paper #1059, Stanford University, Stanford, CA.

Guadagni, Peter M., John D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2** (Summer) 203–238.

Gupta, Sunil. 1988. Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* **XXV** (November) 342–355.

Inman, Jeffrey J., Leigh McAlister, Wayne D. Hoyer. 1990. Promotion signal: Proxy for a price cut. *J. Consumer Res.* **17** (June) 74–81.

——, ——. 1993. A retailer promotion policy model considering promotion signal sensitivity. *Marketing Sci.* **12**(4) 339–356.

Kadiyali, Vrinda, Pradeep Chintagunta, Naufel Vilcassim. 2000. Manufacturer-retailer channel interactions and implications for channel power: An empirical investigation of pricing in a local market. *Marketing Sci.* **19** (Spring) 127–148.

Mason, J. Barry, Morris L. Mayer. 1990. *Modern Retailing: Theory and Practice.* BPI/Irwin, Homewood, IL.

Narasimhan, Chakravarthi, Scott A. Neslin, Subrata K. Sen. 1996. Promotional elasticities and category characteristics. *J. Marketing* **60** (April) 17–30.

Salop, Steven. 1977. The noisy monopolist: Imperfect information, price dispersion, and price discrimination. *Rev. Econom. Stud.* **44** 393–406.

Sawyer, Alan G. 1975. Demand artifacts in laboratory experiments in consumer research. *J. Consumer Res.* **1** (March) 20–30.

Shimp, Terence A., Eva M. Hyatt, David J. Snyder. 1991. A critical appraisal of demand artifacts in consumer research. *J. Consumer Res.* **18** (December) 273–283.

Stiving, Mark, Russell Winer. 1997. An empirical analysis of price endings with scanner data. *J. Consumer Res.* **24** (June) 57–67.

Tellis, Gerard J. 1988. The price elasticity of selective demand: A meta-analysis of econometric models of sales. *J. Marketing Res.* **XXV** (November) 331–341.

Thaler, Richard. 1985. Mental accounting and consumer choice. *Marketing Sci.* **4** (Summer) 199–214.

Villas-Boas, Miguel, Russell Winer. 1996. Endogeneity in brand choice models. Working paper, University of California at Berkeley, Berkeley, CA.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–330.

This paper was received March 20, 2000, and was with the authors 2 months for 2 revisions; processed by Scott Neslin.