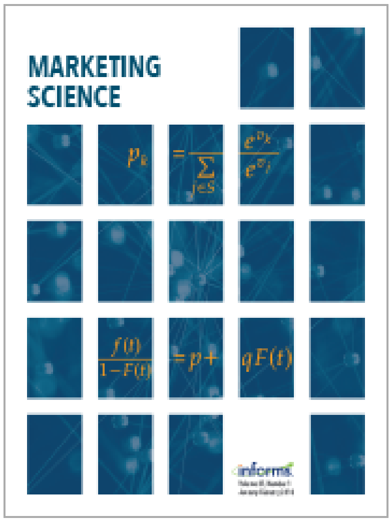

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Research Note—Does Demand Fall When Customers Perceive That Prices Are Unfair? The Case of Premium Pricing for Large Sizes

Eric T. Anderson, Duncan I. Simester,

MARKETING SCIENCE
Vol. 27, No. 3, May–June 2008, pp. 492–500
ISSN 0732-2399 | EISSN 1526-548X | 08 | 2703 | 0492

DOI 10.1287/mksc.1070.0323

Research Note

Does Demand Fall When Customers Perceive That Prices Are Unfair? The Case of Premium Pricing for Large Sizes

Eric T. Anderson
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
eric-anderson@kellogg.northwestern.edu

Duncan I. Simester
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142,
simester@mit.edu

We analyze a large-scale field test conducted with a mail-order catalog firm to investigate how customers react to premium prices for larger sizes of women's apparel. We find that customers who demand large sizes react unfavorably to paying a higher price than customers for small sizes. Further investigation suggests that these consumers perceive that the price premium is unfair. Overall, premium pricing led to a 6% to 8% decrease in gross profits.

Key words: product line pricing; price discrimination; fairness; price promotion; experimental economics
History: This paper was received January 25, 2005, and was with the authors 14 months for 5 revisions; processed by Arvind Rangaswamy. Published online in *Articles in Advance* March 31, 2008.

1. Introduction

Manufacturers of women's apparel often charge retailers higher wholesale prices for large sizes than for smaller sizes of the same items. However, retailers rarely pass the higher wholesale prices onto their customers by charging a premium for the larger sizes. Their reluctance to vary the retail price in response to variations in the wholesale price is apparently driven by concerns that customers will perceive that it is unfair to charge higher prices for larger sizes of the same item.

These concerns about perceived fairness are supported by a growing academic literature describing when fairness concerns will arise and how customers will react. This literature includes theoretical models of fairness together with empirical evidence collected from customer surveys and laboratory experiments. Surprisingly, there have been almost no field studies of the impact of perceived fairness on actual demand. Even retailers themselves have been slow to investigate this issue; most simply choose a policy that they anticipate will reduce fairness concerns without measuring the extent to which these concerns would impact profits.

In this paper we directly measure the impact of perceived fairness on customer demand. In particular, we investigate how customers react when the price of various apparel products is higher for larger sizes than for smaller sizes. We report the findings from a very large-scale field test in which we measured the actual purchasing behavior of real customers. The field test was conducted with a mail-order catalog firm that sells women's fashion clothing in the moderate price range. We designed four different versions of the catalog in which we varied the use of premium pricing for large sizes. Each version was mailed to a random sample of customers and the subsequent demand enabled us to measure the effects of charging a premium for large sizes. The results reveal that customers who demand large sizes react unfavorably to paying a higher price than customers for small sizes. However, this adverse effect is isolated to customers who demand the smallest of the large sizes. The evidence is consistent with these consumers perceiving that the price premium is unfair. Interestingly, there was no corresponding increase in demand among customers purchasing small sizes.

The study was motivated in part by a pilot study in which the catalog company agreed to mail two

Table 1 Study 1 Design and Results

	Prices ($)			Number of units sold			
		Large size premium		Small size		Large size	
Item	Small size	Control	Treatment	Control	Treatment	Control	Treatment
1	129	0	5	8	11	8	6
2	64	0	5	3	3	7	5
3	69	0	5	1	3	2	1
4	49	0	10	4	5	5	4
5	99	0	10	6	1	7	2
6	69	0	10	3	4	1	0
7	79	0	10	5	3	3	3
8	79	0	10	0	1	0	2
Total				30	31	33	23

different versions of a catalog to separate, randomly selected customer samples. At the time of the experiment, the company had recently added two larger sizes to expand its product line. For the new larger sizes, the catalog's manufacturers charged an average wholesale price that was 10% higher than for the four smaller sizes.[1] These wholesale price differences prompted the company to consider charging different retail prices for the different sizes. However, before implementing such a policy the company was interested in learning how customers would react to premium pricing for the larger sizes.

In the pilot study the prices of eight items were manipulated across the two catalog versions. In the control version the prices varied across styles, but were constant across sizes, so that customers paid the same price for a particular style no matter what size they purchased. In the treatment version the prices of the larger-sized items were increased. The prices and demands for the eight items are summarized in Table 1. The catalog, which contained several hundred other items, was mailed to equal numbers of customers in each condition.

The results reveal that demand for small sizes was seemingly unaffected by the price of the larger size. However, demand for large sizes was lower in the treatment version. Although the sample sizes were small, the difference is marginally significant ($p < 0.10$). One interpretation of this finding is simply that the increase in price of the large items made them less attractive. However, the catalog managers were surprised by the extent to which demand had fallen and felt that this was unlikely to be explained solely by own-price elasticity for these items. Instead, there was concern that customers who needed the large sizes were reacting adversely to having to pay more than customers purchasing small sizes. The managers agreed to conduct a follow-up study to investigate this issue. In the follow-up study they agreed to exogenously vary the prices of both sizes so that we could disentangle the two explanations.

[1] The average wholesale prices for the four original sizes (XL, 1X, 2X, 3X) were $15.81, $15.88, $15.80, and $15.91. The small differences in wholesale prices between these four sizes are not statistically different and reflect differences in order volumes and the timing of orders. For the two new larger sizes (4X and 5X) the average wholesale prices were $17.48 and $17.45.

Previous Literature

Within the price fairness literature, there are several streams of literature. We begin with the principle of dual entitlement proposed by Kahneman et al. (1986a, b), to which much of the recent work on price fairness traces its origins. This theory argues that customers' have perceived fairness levels for both firm profits and retail prices. Although firms are entitled to earn a fair profit, customers are also entitled to a fair price. Deviations from a fair price can be justified only by the firm's need to maintain a fair profit. According to this argument, it is fair for retailers to raise the price of snow shovels if the wholesale price increases, but it is not fair to do so if a snowstorm leads to excess demand.

One feature of dual entitlement is that it relies on customers being able to assess what a fair profit is. In the snow shovel example, customers can evaluate fairness against the previous status quo: If the firm was earning fair profits before the snow storm, then raising prices must be exploitative. There is a large research stream focusing on the firm's motivation for changing its prices (see, for example, Bies et al. 1993, Martins and Monroe 1994, Campbell 1999, and Bolton et al. 2003).

More generally, the concept of price fairness is closely related to the reference price literature and yields similar predictions. In the context of this study, the price charged for other sizes of the same item may contribute to the formation of customers' reference prices. The price charged for the other sizes represents an explicit and highly accessible reference price from which fairness judgments can be made. It seems that this reference price interpretation is best interpreted as complementing the price fairness argument, rather than as a substitute or alternative explanation. Both arguments anticipate that customers evaluate the fairness of the transaction against a reference point, and interpret the price charged for other sizes as a determinant of that reference point.

Although there is considerable evidence that customers use reference prices when making purchasing decisions, the question of what information customers use to form a reference price is not fully resolved (Briesch et al. 1997). Several authors have proposed internal or memory-based references (see, for example, Janiszewski and Lichtenstin 1999, Lichtenstin et al. 1988, and Monroe 1990). Indeed, the fairness research focusing on firm's motivations for changing prices (cited above) can be interpreted as indirectly contributing to this literature. Customers interpret the fairness of the current price of snow shovels by comparing it to their recollection of past prices. Other papers have cited external references, focusing primarily on the prices charged by competitors (see, for example: Alba et al. 1994, Buyukkurt 1986, Lichtenstin and Bearden 1989, Dholakia et al. 2005, and Urbany et al. 1988). The evidence in this paper that customers are sensitive to the prices of other sizes of the same item suggests an additional external source that customers use to form their reference prices.

Evidence of price fairness has also been documented in studies of buyer satisfaction in price negotiations. Several studies have shown that buyer satisfaction influences the probability of repeat negotiations and the performance of the negotiated obligations (Barry and Oliver 1996). Although there is an obvious preference for more favorable absolute outcomes (buyers prefer lower prices), Novemsky and Schweitzer (2004) show that buyer satisfaction is also influenced by social comparisons: Was the price higher or lower than the prices negotiated by other customers? Their findings reveal that the seller's profit is less important in explaining buyer satisfaction than the price paid by other buyers. These comparisons with other buyers are particularly influential when there are few other benchmarks available.

Price fairness has also been examined in yield management applications, which typically require that customers pay different prices for the same service (e.g., airlines, hotels, etc.). Anticipation that this practice may lead to customer dissatisfaction, and in turn have an adverse long-term impact on demand, has led to recommendations for mitigating these effects (see, for example, Kimes 2002 and Kambil and Agrawal 2001). Proposals include offering additional services to customers paying higher prices or imposing service restrictions on customers receiving discounted prices.

In a recent paper, Xia et al. (2004) provide an extensive review of the literature related to price fairness, together with a comprehensive model for reconciling the disparate findings. In the course of their review, the authors identify three features of the fairness literature that will be helpful when interpreting the results of this study. First, they argue that fairness is a comparative phenomenon, in which customers evaluate the price that they paid against benchmarks. These benchmarks may include prices at different times, prices paid by different customers, prices of related products, or prices in different stores. They conclude that customers attend to the degree of similarity in the comparisons when evaluating price differences. When transactions are similar, customers will find it more difficult to reconcile price differences, and so are more likely to conclude that a price difference is unfair. Alternatively, if there are important differences between the transactions, customers may attribute price disparities to these differences and fairness concerns are less likely to arise.

Second, Xia et al. (2004, p. 2) highlight an important dichotomy between favorable and unfavorable price comparisons: "...perceived fairness is less severe when the inequality is to the buyer's advantage than when it is to the buyer's disadvantage."[2] They predict that customers will generally not react when price differences are favorable. However, when price comparisons are unfavorable they predict that customers will respond with negative emotions, reduced demand, and negative word of mouth.

Finally, Xia et al. (2004) argue that comparisons with social norms may also help to reinforce fairness perceptions. For example, social norms have been used to explain why customers now rarely complain about price discrimination by airlines (Kimes 2002). The norm in the retailing industry is generally not to charge different prices for different sizes, and the company that participated in this study had not previously charged different prices for different sizes.[3] The departure from this norm in this study may have contributed to the findings. It is possible that premium pricing for larger sizes may not lead to any adverse reactions if it was to become an industry norm.

Structure of the Paper

This paper continues in §2 with a more detailed description of the field study, including a summary of the prices offered in the different versions of the catalog. The results are presented in §3, and we conclude the paper in §4 with a review of the findings.

2. Design of the Study

The study was conducted with the same company that participated in the pilot study. For confidentiality reasons we are unable to identify the name of the catalog. The company sells women's apparel in the plus-size category, which is one of the fastest-growing segments in the apparel industry. The items

[2] As support for this conclusion, Xia et al. (2004) cite Ordónez et al. (2000).

[3] Although certainly not the norm, there are examples of retailers charging premiums for different sizes. L. L. Bean and J. Crew both charge a premium for tall shirts compared to regular shirts. Notice that this policy treats consumers who demand tall shirts equally and treats all consumers who demand regular shirts equally. The distinction between tall and regular sizes may reduce the degree of similarity between the products so that customers interpret this size difference as a sufficient explanation for the price difference. These cost perceptions need not be accurate to be fair—a "small-tall" shirt may have less material and may be less expensive to produce than a "large-regular" shirt.

are all sold under the firm's own private label brand and are only available through the company's catalog. Although clothing with the same brand is not available in retail stores, other companies offer competing brands in both direct and traditional store channels.

The study was conducted in a catalog that contained a total of 155 clothing and accessory products. Many of the products are sold in a range of sizes, and so there are 833 product-size combinations. When the new large sizes were introduced to the product line, the catalog management chose to introduce them with different item numbers. For example, one of the products in the catalog is a wool sweater. For the four smallest sizes the sweater had item number 379235, whereas for the two largest sizes it had item number 635472. Both item numbers are shown in the catalog alongside the range of size options. Across the 155 products there were 211 item numbers.

Catalog managers offered two rationales for the use of multiple item numbers. First, listing the large sizes as separate item numbers alerted customers to the new additions to the product line. Second, not every item was available in the new sizes and managers believed that listing the new sizes as separate item numbers would help to reduce confusion. In its management reports and policies, the company treats different item numbers as different products.

Four different versions of the catalog were designed and each version was mailed to a separate sample of 31,250 customers. Assignment of customers to the different versions was approximately random, with nine-digit zip codes (zip + 4) randomly distributed across conditions and all customers in a specific zip code receiving the same version. The advantage of randomizing by zip code is that it reduces the chance of mailing neighbors catalog versions with different prices. The use of nine-digit zip codes also ensures that the randomization occurs across thousands of zip codes, and so the potential for a confound introduced by the randomization process is small. In practice, the only risk is that the firm implemented the randomization procedure incorrectly. In a prior study conducted with this company, we tested their randomization systems by comparing the past purchasing histories of customers in treatment versus control categories. This comparison confirmed that their randomization procedure was truly random. Unfortunately, we do not have the data to make this comparison in this test. However, we can use the demand for the items that were not manipulated to control for any circulation variations (see later discussion).

The objective of the study was to disentangle the effects of introducing a contrast between the prices of small and large sizes (which we will call the "fairness effect") from the impact of increasing the prices for the large sizes. For this reason, we used the item numbers for different sizes to exogenously vary the prices of large and small sizes across the four conditions. On 38 of the products, the manipulations used the pattern depicted in Table 2.

Under this design, the prices of the large and small sizes were the same on every product in two catalog versions (Versions B and D), whereas a $5 premium was charged for the larger sizes in Versions A and C. By charging different prices for the small sizes in Versions B and D, and in Versions A and C, we also vary the prices of each size separately from the price differences between the sizes. Ideally, the changes in the prices and the price difference would be orthogonal. However, because we cannot change the price difference without also changing the prices, a truly orthogonal design would require charging lower prices on some of the large sizes than on the small sizes (which was not possible). Without a truly orthogonal design, we cannot use a simple univariate approach to interpret the results. Instead, we will use a multivariate approach to estimate the impact of the price difference, while explicitly controlling for the actual price levels.

The $5 premium was chosen to approximate the additional margin that the firm typically needs to earn on larger items in order to maintain its targeted gross profit margin. Of course, for relatively inexpensive items, the $5 premium exceeds the level necessary to achieve this target. For this reason, on 10 of the less expensive products in the catalog a slightly different set of manipulations were used (see Table 3).

Under the guidance of the catalog managers, the values of Y (and, hence, the premiums) ranged from $2 to $4, according to the price of the products. We also varied the prices of 18 products that were only offered in small sizes. This yielded an additional source of price variation to help identify the own-price elasticity independently of the effect of premium pricing. For the remaining 89 products, the prices were held constant across the four versions. In Table 4 we summarize the average price charged for the different product-size combinations in each of the four versions. The average prices match the pattern of manipulations observed in Tables 2 and 3. For example, the prices of the small sizes are approximately $5 larger in Version B than in Version A, whereas the

Table 2 **Price Variation on Expensive Items**

	Version A ($)	Version B ($)	Version C ($)	Version D ($)
Small sizes	X	X + 5	X + 10	X + 10
Large sizes	X + 5	X + 5	X + 15	X + 10

Table 3 Price Variation on Inexpensive Items

	Version A ($)	Version B ($)	Version C ($)	Version D ($)
Small sizes	X	X + Y	X + Y	X + 2Y
Large sizes	X + Y	X + Y	X + 2Y	X + 2Y

prices of the large sizes are the same in these two versions.[4]

There are alternative research designs that may have more efficiently achieved a similar outcome using different experimental conditions. However, varying the content of catalogs is expensive, and preserving the cooperation of a firm restricts discretion over the experimental design. In this case, the final research design reflects extensive interaction between the company and the research team. The final design achieved our research purposes, while satisfying the company's concerns.

The restrictions imposed by the retailer on the experimental design are offset by the many advantages of conducting experimental research in a field setting. First, we can measure the behavior of actual customers engaging in retail transactions in unobtrusive experimental manipulations. This offers greater external validity compared to many laboratory experiments. Second, by exogenously varying the prices shown to different customers, we overcome endogeneity and selection concerns that typically arise when analyzing historical data. Third, the experimental catalog versions can be distributed at the same time to randomly selected customer samples. In contrast, tests conducted in retail stores generally require differences in strategies over time or differences across stores, which introduces the potential for alternative explanations due to intervening events or systematic differences between stores. Fourth, the number and identity of catalog customers who are exposed to the different experimental versions is known. In a retail store it is much more difficult to track the total number and identity of customers who visit a store. Finally, stockouts can distort measurement of demand in a traditional retail setting. For example, there is generally no record of customers who searched for an item and then departed when they could not find it, or customers who were never aware of an item but would have purchased it if it had been on display. In a catalog setting, customers initiate orders in writing or via telephone before learning whether the item is available.

[4] A comparison of Tables 2 and 3 with Table 4 may prompt readers to wonder why the average price of small and large sizes is different in Versions B and D. The small differences in these averages reflect the variation in the prices of the 18 products that were only offered in small sizes.

Table 4 Average Item Price

	Number of items	Version A ($)	Version B ($)	Version C ($)	Version D ($)
Small sizes	245	46.60	51.20	55.51	55.93
Large sizes	114	50.69	50.69	59.50	55.38
Other	474	51.56	52.34	51.56	53.12
Total	833	49.98	53.81	51.78	54.25

3. Results

The catalog company provided data describing the number of units sold for each of the 833 product-size combinations in the four catalog versions. Due to the complexity of the experimental design, we used a multivariate approach to disentangle the contrast between small and large prices from the direct effects of varying prices. Because the dependent variable is a count measure, the multivariate analysis uses Poisson regression.[5] In particular, we assume that the number of units sold for product-size i in catalog version j (Q_{ij}) is drawn from a Poisson distribution with parameter λ_{ij}:

$$\text{Prob}(Q_{ij} = q) = \frac{e^{-\lambda_{ij}}\lambda_{ij}^{q}}{q!}, \quad q = 0, 1, 2, \ldots, \tag{1}$$

where: $\ln(\lambda_{ij}) = \boldsymbol{\beta}\mathbf{X}_{ij}$. The $\mathbf{X}_{ij}$ term denotes the independent variables, whereas $\boldsymbol{\beta}$ denotes the estimated coefficients.

The independent variables include two variables designed to identify the impact of charging different prices for small and large sizes. These variables were defined as:

$$\textit{Price Discount} = (\textit{Large Item Price} - \textit{Small Item Price}) * \textit{Small Item}$$

$$\textit{Price Premium} = (\textit{Large Item Price} - \textit{Small Item Price}) * \textit{Large Item}$$

The *Small Item* and *Large Item* variables are dummy variables indicating whether the item was a small- or large-size. The *Price Discount* variable measures the difference between the prices of the small and large sizes when the item is small and is equal to zero if the item is not small. We will interpret the coefficient estimated for the *Price Discount* variable as a measure of the impact that increasing the difference in prices between the large and small sizes has on demand for small items. The *Price Premium* variable is an analogous measure when the item is large. Note that any price difference between the sizes yields a favorable

[5] We obtained similar results for a linear regression, $Q_{ij} = \boldsymbol{\beta}\mathbf{X}_{ij}$. However, we prefer the Poisson model because it recognizes the discreteness of the dependent variable.

price comparison for the smaller sizes, but an unfavorable comparison for the larger sizes (any price difference results in small items having lower prices than large items).

We emphasize that the labels *Price Discount* and *Price Premium* are adopted merely for expositional convenience. As a reviewer kindly pointed out, the discounts implied by the *Price Discount* term represent discounts that purchasers of small sizes enjoy relative to purchasers of large sizes (not discounts relative to the regular price). Note also that the size of the price difference between the small and large items varies across items. For most of the items the price difference is \$5. However, as previously discussed, for 10 of the relatively less expensive items in the catalog the price differences ranged from \$2 to \$4.

We also included a separate variable to directly control for variance in the own-price of an item (*Price*). By including this own-price measure we can separate the impact of the pure comparison effect from the impact of varying the price of an item. This overcomes the confound that made it difficult to interpret the findings of the pilot study. We also include fixed effects to control for the different products, different sizes, and different catalog versions. The catalog version effects capture the differences (if any) in the distribution of the catalogs due to reliance on randomization by zip code. The 89 items for which prices were not varied across the versions allows us to estimate these version coefficients separately from the price manipulations.

Finally, in previous work conducted in this catalog (Anderson and Simester 2003), we present evidence that \$9 price endings significantly increase demand compared to other price endings. That paper reports findings from Versions B and D of this study (where prices did not vary for large and small sizes). For this reason, we also included a variable identifying when a price ended in \$9. The inclusion of these control variables yielded the following model:

$$\begin{aligned}\mathbf{\beta X}_{ij} &= \beta_1 \mathit{Price\ Discount}_{ij} + \beta_2 \mathit{Price\ Premium}_{ij} \\ &\quad + \beta_3 \mathit{Price}_{ij} + \beta_4 \$9\ \mathit{Ending}_{ij} + \sum \alpha_n \mathit{Product}\ n_{ij} \\ &\quad + \sum \tau_s \mathit{Size}\ s_{ij} + \sum \gamma_j \mathit{Version}\ j_{ij}. \qquad (2)\end{aligned}$$

The α, β, γ, and τ terms are all estimated coefficients. The coefficients of interest are β_1 and β_2: They estimate how contrasts between large and small prices affect demand. Under the Poisson specification the coefficients can be interpreted as percentages, so that β_1 (β_2) estimates the percentage change in demand for small (large) items that results from a \$1 increase in the price difference between small and large sizes.

The findings are reported in Table 5. Although product, size, and version fixed effects are included in our estimation, for ease of exposition we omit them

Table 5 Poisson Regression Results

	Model 1		Model 2		Model 3	
Price discount			−0.009	(0.011)		
Price premium			−0.021	(0.017)		
% Price discount					0.176	(0.433)
% Price premium					−1.232*	(0.745)
Price	−0.017**	(0.006)	−0.017**	(0.006)	−0.014**	(0.006)
\$9 Ending	0.119***	(0.039)	0.118***	(0.040)	0.107***	(0.040)
Sample size	3,332		3,332		3,332	
Log likelihood	−4,813.2		−4,812.3		−4,811.6	

Notes. Item, size, and catalog version effects are omitted from this table. More complete results (including the size and version effects are reported in the appendix in Table A.1. Asymptotic standard errors are in parentheses.

*Significantly different from zero ($p < 0.10$); **significantly different from zero ($p < 0.05$); ***significantly different from zero ($p < 0.01$).

from Table 5 (for completeness we report the size and version coefficients in the appendix, Table A.1). We report three versions of the model. Model 1 represents a benchmark model in which we omit the *Price Premium* and *Price Discount* variables. In Model 2 we include the *Price Premium* and *Price Discount* variables. Behavioral decision theory has shown that in some situations customers are more sensitive to percentage price differences than actual price differences (Kahneman and Tversky 1984, Darke and Freedman 1993, Grewal and Marmorstein 1994). For this reason, we also included a third model in which the *Price Premium* and *Price Discount* variables are calculated using the percentage price differences rather than the absolute price differences.

The *Price Discount* coefficients do not approach significance in Model 2 or Model 3, suggesting that customers who purchase smaller items are not concerned about whether the catalog charges a premium for larger sizes. In contrast, the coefficient for the *Price Premium* coefficient is marginally significant in Model 3 ($p < 0.10$). Recall that this variable measures how the percentage price difference between the large and small sizes affects demand for the large sizes. In particular, the results suggest that charging 1% more for the large size than for the small size will reduce demand for the large size by 1.2%.[6] This finding is a pure contrast effect because we have separately controlled for the price of each item. However, because the significance is marginal, the results offer only weak evidence of a fairness effect.

Recall that Xia et al. (2004) predict that customers will generally not react when price differences are favorable. However, when price comparisons are unfavorable we can expect an adverse impact on demand. The findings in Table 5 are consistent with these predictions. Xia et al. (2004) also conclude that

[6] Because of the marginal significance of this coefficient (and the correspondingly low improvement in the likelihood of Model 3 compared to Model 1), the estimated elasticity in this model should only be used for directional guidance.

Table 6 Prices for Different Sizes of a Wool Sweater

Size	Item number	Price in Version A ($)
XL	379235	53
1X	379235	53
2X	379235	53
3X	379235	53
4X	635472	58
5X	635472	58

customers attend to the degree of similarity in the comparisons when evaluating price differences. When transactions are similar, customers will find it more difficult to reconcile price differences and so are more likely to conclude that a price difference is unfair. Under this reasoning we might expect that customers purchasing the smallest of the large items are most likely to exhibit an unfavorable reaction. This argument is best illustrated with an example.

We earlier described a wool sweater that had item number 379235 for the four smallest sizes and item number 635472 for the two largest sizes (see Table 6). Because customers purchasing the four smaller sizes (XL, 1X, 2X, and 3X) benefit from a favorable price comparison, Xia et al. (2004) predict that we would not observe any fairness effects on demand for these items. The unfavorable price comparisons for the two larger sizes would be consistent with an adverse fairness effect on demand for the 4X and 5X items. Furthermore, notice that the comparison between 3X and 4X prices involves a comparison of more similarly sized items than the comparison between 3X and 5X prices. For this reason, the moderating role played by product similarity argues that any adverse reaction should be stronger for 4X customers than for 5X customers. To investigate these predictions we segmented the large and small sizes more finely into four segments (see Table 7).[7]

We then used this segmentation to modify Equation (2) as follows:

$$\begin{aligned}\boldsymbol{\beta}\mathbf{X}_{ij} = {} & \sum_{k=1}^{2} \beta_k \mathit{Segment}\ k_{ij}\ \mathit{Price\ Discount}_{ij} \\ & + \sum_{k=3}^{4} \beta_k \mathit{Segment}\ k_{ij}\ \mathit{Price\ Premium}_{ij} + \beta_5 \mathit{Price}_{ij} \\ & + \beta_6 \$9\ \mathit{Ending}_{ij} + \sum \alpha_n \mathit{Product}\ n_{ij} \\ & + \sum \tau_s \mathit{Size}\ s_{ij} + \sum \gamma_j \mathit{Version}\ k_{ij}. \qquad (3)\end{aligned}$$

This specification allows us to separately identify how the favorable price comparisons affected customers in Segments 1 and 2 (coefficients β_1 and β_2), and how the unfavorable comparisons affected customers in Segments 3 and 4 (β_3 and β_4).

Table 7 Definition of Segments

Segment	Description	Sizes
Segment 1	Small size	XL, 1X
Segment 2	Small size	2X, 3X
Segment 3	Large size	4X
Segment 4	Large size	5X

The findings are presented in Table 8. The null result for Segments 1 and 2 indicates that these customers did not react to the favorable price comparison. Among customers exposed to an unfavorable price comparison, we only observe an adverse reaction from the customers purchasing the smaller of the two large sizes (Segment 3). These are the customers for whom the size comparisons are most similar, and so for whom the price differences are hardest to reconcile. We conclude that the findings are consistent with the predictions of Xia et al. (2004).

As an alternative explanation, a reviewer suggested that customers for the 4X sizes may have responded to the price premiums by substituting and purchasing the 3X sizes instead. To investigate this possibility, we studied the impact of the price premiums on demand for the 3X sizes alone (recall that Segment 2 measures the joint demand for 2X and 3X sizes). This analysis revealed that the price premiums on the 4X and 5X sizes do not lead to increased demand for 3X sizes, suggesting that substitution cannot explain the drop in demand for the 4X sizes.

To investigate possible aggregation bias, we also estimated several additional models, including specifications that allowed us to compare the impact of the price premiums on different product categories (tops, bottoms, and dresses) and items at different price levels. These analyses revealed a reassuringly

Table 8 Poisson Regression Results

	Model 4		Model 5	
Price discount ∗ Segment 1	−0.016	(0.014)		
Price discount ∗ Segment 2	−0.010	(0.013)		
Price premium ∗ Segment 3	−0.051**	(0.023)		
Price premium ∗ Segment 4	0.007	(0.022)		
% Price discount ∗ Segment 1			−0.311	(0.516)
% Price discount ∗ Segment 2			0.388	(0.490)
% Price premium ∗ Segment 3			−2.090**	(0.996)
% Price premium ∗ Segment 4			−0.461	(0.971)
Price	−0.017***	(0.006)	−0.015**	(0.006)
$9 Ending	0.120***	(0.040)	0.109***	(0.040)
Sample size	3,332		3,332	
Log likelihood	−4,810.0		−4,811.6	

Notes. Item, size, and catalog version effects are omitted from the table. Asymptotic standard errors are in parentheses.

*Significantly different from zero ($p < 0.10$); **significantly different from zero ($p < 0.05$); ***significantly different from zero ($p < 0.01$).

[7] Some items used a finer sizing scheme. Fortunately, there is an equivalent mapping between the schemes: Segment 1 included sizes 14–20, Segment 2 included sizes 22–26, Segment 3 included sizes 28–30, and Segment 4 included sizes 32–34.

similar pattern of findings across the different subgroups of items: The *Price Premium* variable is negative, and the effect is concentrated in the demand for the 4X sizes. However, this segmentation led to a loss of statistical power and a corresponding reduction in statistical significance.

We earlier recognized that the rather unusual nature of the price differences in this study may have contributed to the findings. Because charging different prices for different sizes of fashion apparel is not a norm in retailing, the comparisons with social norms may have contributed to customers' perceptions of unfairness. An additional feature of the study that may have contributed to these results is that the distinction between small and large sizes is somewhat arbitrary. This distinction leads to the prices following a step function across the size range, which may have made it more difficult to reconcile the price premium for large sizes. The findings may be different if we had been able to test different prices across each of the sizes.

In other findings of interest, the *Price* and *$9 Ending* coefficients are both statistically significant in all four models. Reassuringly, their coefficients do not vary significantly across the four model specifications. We also reestimated the models without the *$9 Ending* variable, and the pattern of results did not change.

3.1. Managerial Implications

The negative coefficients for the *Price* and *Price Premium* variables together indicate that raising the price of the larger sizes will decrease demand for two reasons: The price increase makes the product less attractive and introduces an unfavorable price comparison with the smaller sizes. This does not itself imply that the premiums on larger sizes are unprofitable, because these demand effects are offset by the additional margin on the large sizes. Whether this additional margin compensates for the loss in demand is an empirical question.

To further investigate this issue, we used the price and cost-of-goods-sold data together with the estimated demand from each item to calculate the gross profits earned from each version. We calculated the gross profit of item i in catalog version j as: $\pi_{ij} = Q_{ij}(p_{ij} - c_i)$. Recall that Q_{ij} denotes the number of units of item i sold in catalog version j. We use the estimated quantity from Model 5 and remove the version effects, which allows us to eliminate any confounds due to differences in distribution. The p_{ij} term denotes the price of item i in version j, whereas c_i denotes the wholesale price of each item (which did not vary across catalog versions). We then aggregated across items to calculate the total gross profit earned from each version: $\pi_j = \sum_i Q_{ij}(p_{ij} - c_i)$. The resulting gross profits are summarized in Table 9 and confirm that charging a premium for large sizes was not profitable.

Table 9 Gross Profit Earned from Each Catalog Version

		All items ($)		Large items ($)	
Catalog version	Price premium for large size?	Total profit	Total revenue	Total profit	Total revenue
Version A	Yes	45,519	71,916	4,538	7,025
Version B	No	49,354	74,320	7,513	10,723
Version C	Yes	46,590	71,792	5,362	8,248
Version D	No	49,826	74,909	7,332	10,748

4. Conclusions

We have reported the findings from a large-scale field test that allowed us to investigate how customers react when a firm charges different prices for different sizes of the same apparel item. Using four experimental versions of a catalog, we exogenously varied prices of different sizes of a sample of items. This design made it possible to disentangle the effects of introducing a contrast between the prices of small and large sizes from the impact of increasing the prices for the large sizes. The findings confirm several conclusions from Xia et al. (2004) in their recent survey of the price fairness literature. Customers react adversely when faced with an unfavorable price comparison, and this adverse reaction is strongest among customers for whom the price comparison is most similar. In contrast, we did not observe any reaction among customers who were presented with a favorable price comparison.

Our estimates show that the magnitude of the price fairness effect is twice as large as the direct effect of raising price. For example, the pure-price effect of 10% price premium for large sizes is an 8% loss in demand, accompanied by an additional 20% loss in demand due to the price fairness effect. This finding is consistent with a growing body of literature showing that the direct effect of varying prices is often smaller than the effect of varying price cues, such as sale signs or reference prices.

Previous work is comprised primarily of perceptual measures collected in a laboratory setting. In contrast, our findings rely upon demand measures from actual customers engaged in real transactions. These findings serve to validate the previous work in a new experimental setting. However, we caution that our measures are both a strength and limitation of the study. Although they offer an accurate measure of demand, we have interpreted the demand effects as evidence of fairness concerns, yet we have no direct measures of perceived fairness. In the absence of perceptual measures, we can only point to the consistency with the previous literature on price fairness to support this interpretation. In this respect, the findings in this study both complement the previous literature and benefit from that literature as a basis for interpretation.

A limitation of the study is the lack of orthogonality in the experimental design. As we discussed when introducing the experimental design, this is unavoidable because increasing the price difference requires increasing the price for the large size (relative to the small size). As a result, variations in the price premium cannot be independent of variations in the price of both the large and small sizes. The main cost of this limitation is that it introduces a potential confound in our analysis of the impact of price premium. As a result, we cannot use a univariate approach to analyze the results, and instead must use a multivariate approach to disentangle the different effects.

Appendix

Table A.1 Poisson Regression Results

	Model 1		Model 2		Model 3	
	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Control	−0.360	(0.429)	−0.360	(0.441)	−0.486	(0.446)
Version A	−0.256	(0.445)	−0.259	(0.458)	−0.397	(0.463)
Version B	−0.174	(0.436)	−0.201	(0.447)	−0.315	(0.450)
Version C	−0.193	(0.445)	−0.222	(0.457)	−0.339	(0.460)
Size 14	0.319	(0.489)	0.308	(0.494)	0.226	(0.496)
Size 16	0.838*	(0.485)	0.828*	(0.490)	0.745	(0.491)
Size 18	1.031**	(0.484)	1.020**	(0.489)	0.938*	(0.490)
Size 20	0.811*	(0.485)	0.800	(0.490)	0.718	(0.491)
Size 22	0.670	(0.485)	0.660	(0.490)	0.577	(0.491)
Size 24	0.475	(0.486)	0.464	(0.491)	0.382	(0.492)
Size 26	0.384	(0.487)	0.374	(0.492)	0.292	(0.493)
Size 28	0.698	(0.503)	0.695	(0.504)	0.670	(0.504)
Size 30	0.608	(0.520)	0.605	(0.520)	0.580	(0.520)
Size 32	0.441	(0.509)	0.437	(0.509)	0.413	(0.509)
Size 34	1.097**	(0.496)	1.094**	(0.497)	1.069**	(0.497)
Size XL	−0.517	(0.451)	−0.517	(0.451)	−0.517	(0.451)
Size X1	−0.017	(0.449)	−0.017	(0.449)	−0.017	(0.449)
Size X2	0.037	(0.449)	0.037	(0.449)	0.037	(0.449)
Size X3	−0.234	(0.450)	−0.234	(0.450)	−0.234	(0.450)
Size X4	0.035	(0.077)	0.035	(0.077)	0.035	(0.077)
Price discount			−0.009	(0.011)		
Price premium			−0.021	(0.017)		
% Price discount					0.176	(0.433)
% Price premium					−1.232*	(0.745)
Price	−0.017**	(0.006)	−0.017**	(0.006)	−0.014**	(0.006)
$9 Ending	0.119***	(0.039)	0.118***	(0.040)	0.107***	(0.040)
Sample size	3,332		3,332		3,332	
Log likelihood	−4,813.2		−4,812.3		−4,811.6	

Notes. Item effects are omitted from the table.

*Significantly different from zero ($p < 0.10$); **significantly different from zero ($p < 0.05$); ***significantly different from zero ($p < 0.01$).

References

Alba, J. W., S. Broniarczyk, T. Shimp, J. E. Urbany. 1994. The influence of prior beliefs, frequency cues, and magnitude cues on consumers' perceptions of comparative price data. *J. Consumer Res.* **21** 219–235.

Anderson, E. T., D. I. Simester. 2003. Effects of $9 price endings on retail sales: Evidence from field experiments. *Quant. Marketing Econom.* **1**(1) 93–110.

Barry, B., R. Oliver. 1996. Affect in dyadic negotiation: A model and propositions. *Organ. Behav. Human Decision Processes* **67** 127–143.

Bies, R. J., T. M. Tripp, M. A. Neale. 1993. Procedural fairness and profit seeking: The perceived legitimacy of market exploitation. *J. Behavioral Decision Making* **6**(December) 243–256.

Bolton, L., L. Warlop, J. W. Alba. 2003. Consumer perceptions of price (un)fairness. *J. Consumer Res.* **29**(March) 474–491.

Briesch, R., L. Krishnamurthi, T. Mazumdar, S. P. Raj. 1997. A comparative analysis of reference price models. *J. Consumer Res.* **24**(2) 202–214.

Buyukkurt, B. K. 1986. Integration of serially sampled price information: Modeling and some findings. *J. Consumer Res.* **13** 357–373.

Campbell, M. C. 1999. Perceptions of price unfairness. *J. Marketing Res.* **36**(May) 187–199.

Darke, P. R., J. L. Freedman. 1993. Deciding whether to seek a bargain: Effects of both amount and percentage off. *J. Appl. Psych.* **78**(December) 960–965.

Dholakia, U. M., J. H. Jones, I. Simonson. 2005. The effect of explicit reference points on consumer choice and online bidding behavior. *Marketing Sci.* **24**(2) 194–205.

Grewal, D., H. Marmorstein. 1994. Market price variation, perceived price variation, and consumers' price search decisions for durable goods. *J. Consumer Res.* **21**(December) 453–460.

Janiszewski, C., D. Lichtenstin. 1999. A range theory account of price perception. *J. Consumer Res.* **25** 353–368.

Kahneman, D., A. Tversky. 1984. Choices, values and frames. *Amer. Psychologist* **39** 341–350.

Kahneman, D., J. L. Knetsch, R. H. Thaler. 1986a. Fairness and the assumptions of economics. *J. Bus.* **59**(October) S285–S300.

Kahneman, D., J. L. Knetsch, R. H. Thaler. 1986b. Fairness as a constraint on profit seeking: Entitlements in the market. *Amer. Econom. Rev.* **76**(September) 728–741.

Kambil, A., V. Agrawal. 2001. The new realities of dynamic pricing. *Outlook* **2** 15–21.

Kimes, S. 2002. Perceived fairness of yield management. *Cornell Hotel Restaurant Admin. Quart.* **43**(1) 21–30.

Lichtenstein, D. R., W. O. Bearden. 1989. Contextual influence on perceptions of merchant-supplied reference prices. *J. Consumer Res.* **16** 55–66.

Lichtenstein, D. R., P. H. Bloch, W. C. Black. 1988. Correlates of price acceptability. *J. Consumer Res.* **15** 243–252.

Martins, M., K. B. Monroe. 1994. Perceived price fairness: A new look at an old construct. C. T. Allen, D. R. John, eds. *Advances in Consumer Research*, Vol. 21. Association for Consumer Research, Provo, UT, 75–78.

Monroe, K. B. 1990. *Pricing: Making Profitable Decisions*. McGraw-Hill/Irwin, New York.

Novemsky, N., M. E. Schweitzer. 2004. What makes negotiators happy? The differential effects of internal and external social comparisons on negotiator satisfaction. *Organ. Behav. Human Decision Processes*. **95**(2) 186–197.

Ordónez, L. D., T. Connolly, R. Coughlan. 2000. Multiple reference points in satisfaction and fairness assessment. *J. Behavioral Decision-Making* **13**(3) 329–344.

Urbany, J. E., W. O. Bearden, D. C. Weilbaker. 1988. The effect of plausible and exaggerated reference prices on consumer perceptions and price search. *J. Consumer Res.* **15** 95–110.

Xia, L., K. B. Monroe, J. L. Cox. 2004. The price is unfair! A conceptual framework of price fairness perceptions. *J. Marketing* **68**(4) 1–15.