Private Label vs. National Brand Price Sensitivity:
Evaluating Non-experimental Identification Strategies

Nathan M. Fong, Duncan I. Simester, and Eric T. Anderson

May 17, 2011

Abstract

Predictions vary on how the price sensitivity of private labels differs from that of other products. Using a large-scale field experiment we first show that private label demand is less price sensitive than national brand demand. We then compare the estimates from the experimental study with estimates using the same retailer’s transaction history. This allows us to evaluate several methods for controlling for the endogeneity of prices in non-experimental studies. Measuring price sensitivity in the historical data without accounting for endogeneity performs poorly. Instrumental variables estimates with commodity prices as instruments and regression discontinuity estimates also differ from the experimental benchmark. However, estimates using wholesale prices as an instrument closely replicate the experimental estimates. These findings indicate that the wholesale price is an effective instrument for retail price.

*Nathan M. Fong is a doctoral student (e-mail: nfong@sloan.mit.edu) and Duncan I. Simester is NTU Professor of Management Science (e-mail: simester@mit.edu) at the MIT Sloan School of Management. Eric T. Anderson is Hart-marx Professor of Marketing (e-mail: eric-anderson@northwestern.edu) at the Kellogg Graduate School of Management, Northwestern University. This article has benefitted from comments by seminar participants at MIT, Rotterdam School of Management, Temple University, Tilburg University, University of Wisconsin - Madison, Yahoo! Research, and the 2010 Marketing Science conference.
1 Introduction

Private label products feature prominently in contemporary retailing. Shares of total sales have grown substantially both in the U.S. and internationally over recent years, and private labels are perceived as a strong opportunity for retailer growth. Understanding the price sensitivity of demand for private labels is critical for determining pricing strategy and managing a private label portfolio. However, it is uncertain, a priori, whether we expect to find that private labels are more or less price sensitive than national brands. Using a large-scale field experiment, we find that price sensitivity for private labels is lower than that of national brands. The experimental results serve as a benchmark to evaluate non-experimental (or ‘observational’) estimates based on the same retailer’s transaction history. We find that estimates vary substantially depending on the approach to estimation. Instrumental variables estimation using wholesale prices was most effective in replicating the experiment.

The field experiment was conducted through a retail chain. Regular prices were exogenously manipulated for a sample of 192 items across 18 retail locations, and maintained for a 17 week period. The experiment provides strong internal validity. However, field experiments are expensive to run, and are often unavailable to researchers and managers. The availability of the firm’s historical data allows us to generalize the finding to a broader sample of products, and in the process evaluate non-experimental techniques based on assumptions that are otherwise difficult to validate.

Using the historical data, we estimate price elasticities using instrumental variables (IV) to account for endogeneity of prices. This approach requires an instrument that satisfies two conditions: it must be correlated with prices, and uncorrelated with the statistical error (the exclusion restriction). The proposed instruments include wholesale prices and commodity price indexes. A qualitative assessment suggests that while wholesale prices are more strongly correlated with retail prices, the commodity prices are more likely to satisfy the exclusion restriction. Using either instrument, IV confirms the finding that private label demand is less price sensitive than national brands, whereas ordinary least squares (OLS) estimates would not have found a significant difference. Standard diagnostics offer little guidance on which set of instruments is better. However, the wholesale price turns out to be remarkably effective at replicating the experimental estimates.

We also apply a regression discontinuity (RD) analysis to the historical data. The retailer’s prices are characterized by infrequent and discrete changes, so that they resemble a step function over time. These changes are likely to be correlated with unobserved demand shocks, but the unobserved factors may be relatively stable around the time of the price changes. Under this assumption, a comparison of sales immediately before and after price changes can be used to estimate price sensitivity. The

---

1 In 2010, private labels held a 17.4% share in the U.S., at least a 2% increase over 5 years, according to Nielson (2011).
experimental benchmark reveals that RD is not a valid approach to measuring price sensitivity in this setting. The estimates are larger in magnitude than the experimental estimates due to promotional activity coinciding with price changes. Moreover, the difference between private label and national brand elasticity estimates is attenuated due to the timing of price changes.

Field experiments have previously been used as a means of validating non-experimental methods, primarily in the context of employment and education program evaluation. The methods have included sample selection, matching, and regression discontinuity. To our knowledge, this study offers the first experimental validation of a conventional instrumental variables estimator, and the first such validation of estimation strategies for price sensitivity. The importance of these results is amplified by the widespread use of instrumental variables techniques and the recognition that price is a potentially endogenous variable.

Section 2 reviews the related literature, including alternative predictions for why we might expect differences in private label and national brand price sensitivities. Section 3 presents the results from the field experiment. Section 4 presents the instrumental variable estimates and Section 5 presents the regression discontinuity estimates. Section 6 concludes.

2 Predictions

The relationship between private labels and national brands is complex, and figures as a familiar topic in the marketing literature (Sethuraman, 2009, offers a recent review). Intuitively appealing theories can yield different predictions about how price sensitivity differs between these classes of products. While not an exhaustive catalog, this section describes several hypotheses supported by the marketing literature.

2.1 Higher price sensitivity for private labels

Customer differences may affect price sensitivity for private labels relative to national brands. Hansen, Singh, and Chintagunta (2006) report that household tendency to buy store brands is correlated across categories, and is also positively correlated with price sensitivity. Looking at variation across locations, Hoch (1996) reports that private labels are more prevalent (i.e. have higher shares) in stores with more price-sensitive consumers, as measured by store-level price elasticity, and demographics associated with price sensitivity (e.g. age and income; see Hoch, Kim, Montgomery, and Rossi, 1995 for details). If the customers that tend to buy private labels are more price sensitive, then we might expect private label demand to be more price sensitive than national brand demand.

Another reason why private label demand might be more price sensitive is because private label
products lack strong differentiation compared to national brands. Scott Morton and Zettelmeyer (2004) argue that retailers position their private labels to imitate leading national brands for strategic reasons: they can improve their negotiating power (also see Narasimhan and Wilcox, 1998), and free-ride on marketing activities. Empirically, they find that private labels often match national brands in appearance, and are placed next to national brands on store shelves. If this signals that the private label is equivalent to the national brand, it should result in higher price sensitivity. While this strategy should also increase the price sensitivity of targeted national brands, the effects are unlikely to be symmetric because we expect national brands to retain some loyal customers.

While a retailer may attempt to imitate the positioning of the national brand, customers do not necessarily treat the private label and national brand as undifferentiated. This implies a difference in brand credibility. Erdem, Swait, and Louviere (2002) show that high brand credibility lowers price sensitivity, in part because it lowers consumers’ perceived risk. Thus, we might expect brand credibility to matter less in low risk categories, and to matter more in risky product categories. Studies have found that perceived risk in a product category affects private label success (Bettman, 1974; Narasimhan and Wilcox, 1998). Together, these findings suggest that private labels have less brand credibility and higher price sensitivity.

Differences in market share and scaling effects may contribute to differences in price sensitivity. All else equal, a smaller market share will increase measures of price sensitivity scaled to a product’s own sales volume. This includes the most commonly used measure, price elasticity, because it affects the denominator of the elasticity. If private label shares tend to be smaller than those of leading national brands, all else equal we might find higher price elasticities. Similar reasoning has been offered as an explanation for asymmetries in cross-price elasticities (Sethuraman, Srinivasan, and Kim, 1999).

2.2 Lower price sensitivity for private labels

We also offer several reasons why private labels could be less price sensitive than national brands. One possibility is that retailers are more likely to launch private labels in categories where they expect inelastic demand. Raju, Sethuraman, and Dhar (1995) find that store brand entry and store brand shares are correlated with low national brand price sensitivity, because the retailer can already earn higher profits when manufacturers are in price competition with each other. This is consistent with Scott Morton and Zettelmeyer’s (2004) finding that retailers are more likely to introduce store brands against powerful national brands. As a result, private labels could be more common in categories with low price sensitivity.

Contrary to the prediction that private label products lack differentiation, they may derive
differentiation from the retailers that market them. Between retailers, there may be more intra-brand competition for national brands than for private labels, which are harder to directly compare (Steiner, 2004). A given retailer may choose not to carry competing products at a similar price point. If the private label has the lowest price in a category, a price-sensitive customer’s next-best alternative may be to exit the category. Previous research suggests that category choice is less sensitive to marketing actions than brand choice (typically found in the decomposition of promotional elasticities, e.g. Gupta, 1988, and Bell, Chiang, and Padmanabhan, 1999), and a similar effect could extend to regular prices. This outcome could also result from non-compensatory preferences where a segment screens on price: a preference for the cheapest option would result in lower price sensitivity for private labels. Thus, a lack of in-store substitutes could result in lower price sensitivity for private labels.

The information available to customers may also affect price sensitivity. If customers have poor price knowledge for private labels relative to national brands, they may not recognize price changes. As a result, they may be less sensitive to posted prices. Price knowledge can vary for a variety of reasons (for examples, see Vanhuele and Drèze, 2002). Price knowledge may be lower for private labels because they are harder to compare across retailers, and customers may have less exposure to private label products. Hoch and Lodish (1998) report survey results showing worse price recall for private labels.

When they are unable or unmotivated to compare prices, customers tend to rely on price cues (Inman, McAlister, and Hoyer, 1990). Anderson, Cho, Harlam, and Simester (2010) find that customer price knowledge has a positive relationship with price sensitivity, and a negative relationship with price cue sensitivity. Thus, lower price knowledge for private labels could result in lower price sensitivity. Furthermore, if customers perceive private labels as a signal of low prices, they may be less attentive to private label prices, reinforcing the lower price knowledge for private labels.

<table>
<thead>
<tr>
<th>Table 1: Summary of Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Higher price sensitivity for private labels</strong></td>
</tr>
<tr>
<td>Price sensitive customers select into private labels</td>
</tr>
<tr>
<td>Private labels are undifferentiated</td>
</tr>
<tr>
<td>Private labels lack brand credibility, perceived as high risk</td>
</tr>
<tr>
<td>Scaling effects for certain measures of price sensitivity</td>
</tr>
<tr>
<td><strong>Lower price sensitivity for private labels</strong></td>
</tr>
<tr>
<td>Selective entry by retailers</td>
</tr>
<tr>
<td>Lack of substitutes at lowest price point</td>
</tr>
<tr>
<td>Lower price knowledge for private labels</td>
</tr>
</tbody>
</table>
2.3 Previous findings

In related research, Hoch and Lodish (1998) studied the effect of the price gap between private labels and national brands using a field experiment. They vary the price gap by changing all the private label prices in the analgesics category. Private label sales are relatively insensitive to the size of the gap. However, this measures sensitivity to the category pricing policy, and does not measure the effect of national brand prices. Category elasticities are typically lower than item-level elasticities, which could account for the price insensitivity. Also, customers may react differently if the change in the price gap is due to a change in national brand prices. In the experiment reported in this paper, prices vary on individual items for both private labels and national brands.

Non-experimental findings on the price sensitivity of private labels are inconclusive. In a cross-category study, Narasimhan, Neslin, and Sen (1996) find that private label share is unrelated to promotional elasticities across 108 categories. Sivakumar and Raj (1997) decompose price effects on brand choice and category choice. They find that private labels are less sensitive to price decreases, and they find mixed results for price increases. This analysis is limited to four categories, and the source of the variation in prices is mainly due to price promotions. In this paper we examine the effects of regular prices. Cotterill, Putsis, and Dhar (2000) examine regular price effects, pooling across a broad sample of products, and estimate a private label elasticity that was only slightly lower than national brand elasticity. However, the data is aggregated to the category level, and is based on differences across categories and markets. We expect to find sharper differences, based on variation over time for individual products.

In a meta-analysis of 1,851 price elasticity estimates (147 of which are for private labels) from 81 publications, Bijmolt, Van Heerde, and Pieters (2005) report no significant effect of brand ownership on price elasticity (with average elasticities of $-2.67$ for manufacturer brands and $-2.59$ for private labels). While the meta-analysis combines estimates from diverse studies, it attempts to control for differences in retail setting and methodology. The endogeneity of marketing decision variables may contribute to this null result: only 293 of the 1,851 estimates accounted for endogeneity of prices. Table 2 reviews a sample of elasticity estimates from the literature that were not included in the meta-analysis. Though this table is limited to a few categories, we observe a tendency for the private label elasticity estimates to be lower than those of national brands. These more recent studies are also more likely to account for endogeneity.

Supply and demand are a classic example of simultaneity, meaning that retailer and customer decisions affect each other. When estimating the price elasticity of demand, it is important to take this simultaneity into account. If retailers use information about demand to set prices, but we do not incorporate this information into our model, we may attribute changes in demand to changes
Table 2: Published Estimates of Private Label and National Brand Elasticities

<table>
<thead>
<tr>
<th>Authors</th>
<th>Journal</th>
<th>Year</th>
<th>Model</th>
<th>Instruments</th>
<th>Category</th>
<th>PL</th>
<th>NB Avg.</th>
<th>NB #</th>
<th>Avg. Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen and Yang</td>
<td>JMR</td>
<td>2007</td>
<td>Choice</td>
<td>OJ</td>
<td>-3.18</td>
<td>-4.54</td>
<td>2</td>
<td></td>
<td>1.36</td>
</tr>
<tr>
<td>Chintagunta</td>
<td>JMR</td>
<td>2002</td>
<td>Choice</td>
<td>Wholesale Prices</td>
<td>-1.81</td>
<td>-2.65</td>
<td>4</td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>Chintagunta, Bonfrer, and Song</td>
<td>MS</td>
<td>2002</td>
<td>Choice</td>
<td>Commodity Prices</td>
<td>-2.24</td>
<td>-2.79</td>
<td>1</td>
<td></td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Frozen Pasta</td>
<td>-3.85</td>
<td>4</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>Cotterill, Putsis, and Dhar</td>
<td>JB</td>
<td>2000</td>
<td>LA/AIDS</td>
<td>Item Size, Industry Concentration</td>
<td>-0.98</td>
<td>-1.07</td>
<td>1</td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Milk</td>
<td>-0.92</td>
<td>-2.05</td>
<td>1</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Butter</td>
<td>-2.49</td>
<td>-1.50</td>
<td>1</td>
<td>-0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bread</td>
<td>-0.80</td>
<td>-1.31</td>
<td>1</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pasta</td>
<td>-3.66</td>
<td>-1.46</td>
<td>1</td>
<td>-2.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Margarine</td>
<td>-6.38</td>
<td>-1.39</td>
<td>1</td>
<td>-4.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Instant Coffee</td>
<td>-0.37</td>
<td>-1.03</td>
<td>1</td>
<td>0.66</td>
</tr>
<tr>
<td>Kamakura and Kang</td>
<td>JR</td>
<td>2007</td>
<td>Log-Log</td>
<td>Toothpaste</td>
<td>-2.20</td>
<td>-2.10</td>
<td>8</td>
<td></td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Toothbrushes</td>
<td>-0.29</td>
<td>-1.81</td>
<td>7</td>
<td></td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Apple Juice</td>
<td>-8.00</td>
<td>-3.61</td>
<td>3</td>
<td>-4.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cranberry Blends</td>
<td>-3.88</td>
<td>-5.01</td>
<td>4</td>
<td>1.13</td>
</tr>
<tr>
<td>Song and Chintagunta</td>
<td>MS</td>
<td>2006</td>
<td>Choice</td>
<td>Wholesale Prices</td>
<td>-3.51</td>
<td>-3.94</td>
<td>3</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liquid Detergent</td>
<td>-1.56</td>
<td>-1.94</td>
<td>3</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liquid Softener</td>
<td>-2.19</td>
<td>-2.88</td>
<td>3</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Softener Sheets</td>
<td>-1.18</td>
<td>-4.70</td>
<td>3</td>
<td>3.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liquid Detergent</td>
<td>-1.50</td>
<td>-2.51</td>
<td>3</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liquid Softener</td>
<td>-4.39</td>
<td>-2.40</td>
<td>3</td>
<td>-1.99</td>
</tr>
</tbody>
</table>

Note: Table includes elasticity estimates that were not included in the Bijmolt, Van Heerde, and Pieters (2005) meta-analysis. Reported values are simple averages by category.

in price when they have actually been caused by other factors. As an example, in the spring of 2010, the municipal water supplies for many cities in metro Boston were briefly declared unsafe for drinking following a water main break. As a result, the quantity demanded for bottled water increased in many communities. A few retailers responded to this upwards shift in demand by increasing prices; however, it would be mistaken to attribute the increase in demand to the increase in price (Ailworth, 2010). Any source of correlation between prices and the statistical error term can produce inconsistent estimates, including simultaneity, omitted variables, and measurement error.

The goal of this paper is to provide a robust comparison of consumer price sensitivity for private label and national brand products based on causal inferences about price effects. One way to address the endogeneity of prices is through the use of an experiment, but the scope of such an experiment is necessarily limited. To generalize the findings to a broader sample, and show how endogeneity might contribute to the mixed findings in the literature, it is useful to reconcile the experimental findings with estimates based on the historical transaction data. Thus, we report both experimental and non-experimental approaches to this problem.

The use of an experimental benchmark as a means of evaluating non-experimental estimates has a rich history in program evaluation. A landmark example is LaLonde (1986), who compares experimental and non-experimental estimates of job training outcomes to demonstrate the difficulty in recovering the experimental estimates using non-experimental methods. This initiated a stream of research that has continued to improve non-experimental program evaluation (e.g. Heckman, Ichimura, and Todd, 1997; Dehejia and Wahba 1999). In an example from education research, Aiken, West, Schwalm, Carroll, and Hsiung (1998) obtained similar results evaluating a program using a randomized experiment and a regression discontinuity design. In the regression discontinuity version of the study, assignment to a program was based on cutoff values of standardized test scores. However, validation in the context of education does not imply that regression discontinuity will work in marketing applications.

In general, there is a shortage of research using field experiments to validate conventional instrumental variables estimators. The closest previous research uses experiments to validate structural models. Todd and Wolpin (2006) use a school subsidy experiment to validate the predictions of a structural model of schooling and fertility decisions. While the estimation of the model requires the specification of exclusion restrictions, the experiment is used to evaluate the overall predictive ability of the economic model, and not the exclusion restriction in isolation. In comparison, we focus on validating the proposed instruments using less restrictive specifications.

2 By ‘conventional,’ we refer broadly to instrumental variables estimators other than randomized experiments and regression discontinuity designs, either of which can be framed as a special case of instrumental variables.
3 Field Experiment

This section presents estimates of private label and national brand price elasticities based on the results of a large scale field experiment. Experiments are an ideal way to measure the effect of endogenous variables, as changing one variable at a time controls for omitted variables and other sources of bias. Compared to a laboratory experiment, field experiments have the advantage of taking place in a natural setting, and the participants are not aware that an experiment is taking place. The customers of the stores hosting the experiment make real decisions as they normally would, ensuring incentive compatibility.

Field experiments have become increasingly common in economics and marketing research, and have been used to study price sensitivity on several occasions. An early example is Curhan (1974), in which prices and other marketing variables were varied for supermarket produce. Hoch and Lodish (1998) and Hoch, Drèze, and Purk (1994) both varied supermarket category-level prices, to compare broad pricing strategies.

3.1 Experimental Design

The experiment was conducted with the cooperation of a large retail chain. The stores sell a typical array of products in the grocery, health and beauty and general merchandise categories. The stores are smaller than most supermarkets and are located in convenient residential and urban locations. The experiment took place at 18 stores located in a single metropolitan area. The test items were sampled in a hierarchical manner: a single item was initially selected at random to represent each category, and then the final selections were reviewed to avoid including close substitutes or complements. Of the 192 items included, 47 were private labels. The items and stores were randomized into blocks, and the experimental conditions were rotated through these blocks to ensure a balanced design. A description of the experiment and its results were also reported in Anderson, Cho, Harlam, and Simester (2010).

The experiment included two pricing conditions: a control condition, in which the regular prices were maintained, and a discount condition, in which the regular prices were reduced by 12%. There was no unusual signage or other indication that the price had been changed in the discount condition. Temporary promotions continued to affect prices in both conditions. This arrangement was maintained for a 17 week period. The outcome of the experiment was recorded as the units sold and average prices paid for each item, at each store, for each week of the experiment, resulting in a sample of 39,168 observations.

\[^3\text{There was an additional experimental condition that varied the presence of price cues. This is used by Anderson, Cho, Harlam, and Simester (2010) to compare the effect of price knowledge on responses to price changes and price cues.}\]
The 12% reduction in prices resulted in a 13% average increase in quantities. There was a 15% increase for national brands, and an 8% increase for private labels, averaged at the item level. The corresponding average elasticities are $-1.26$ for national brands and $-0.68$ for private labels. Pooling over all items, the difference in units sold (between conditions) is significantly smaller for private labels than for national brands, suggesting that the private labels are less sensitive to the discount (see univariate results in Table 3).

### Table 3: Field Experiment Average Weekly Units by Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Discount</th>
<th>Control</th>
<th>Difference</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>10.58**</td>
<td>8.84**</td>
<td>1.74**</td>
<td>2,465</td>
</tr>
<tr>
<td>Brand</td>
<td>(0.79)</td>
<td>(0.58)</td>
<td>(0.27)</td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>5.41**</td>
<td>5.22**</td>
<td>0.19†</td>
<td>799</td>
</tr>
<tr>
<td>Label</td>
<td>(0.42)</td>
<td>(0.43)</td>
<td>(0.10)</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td>1.55**</td>
<td>(0.48)</td>
</tr>
</tbody>
</table>

† Significantly different from zero, $p < 0.10$
* Significantly different from zero, $p < 0.05$
** Significantly different from zero, $p < 0.01$

Standard errors in parentheses.

#### 3.2 Poisson Regression

Typically, the weekly quantity for a given item in a store is fewer than ten units. The number of units sold is a highly skewed, nonnegative ‘count’ measure, and there are many instances of zero units sold in a period (occurring for 20% of item-store-week combinations). We incorporate these natural restrictions in a multivariate analysis using Poisson regression.

We will use the resulting estimates, an average elasticity that accounts for item and store differences, as a benchmark to evaluate the non-experimental results.

The Poisson regression assumes that the number of units sold, $Q_{ist}$ (for item $i$, in store $s$, in week $t$), follows a Poisson distribution with rate parameter $\lambda_{ist}$:

$$\Pr [Q_{ist} = q] = \frac{e^{-\lambda_{ist}}\lambda_{ist}^q}{q!}, \; q = 0, 1, 2, \ldots$$  

---

Parameter estimates for Poisson regression are known to be consistent when the conditional mean of $Q_{ist}$ is correctly specified; however, the maximum likelihood standard errors depend on the restrictive Poisson distribution assumption. We report quasi-maximum likelihood (QML) standard errors, which relax the Poisson assumption by leaving the variance of $Q_{ist}$ unspecified (Wooldridge, 1999). Standard errors were clustered on item, which allows for serial correlation over time and across stores.
The Poisson parameter, which is the mean (as well as the variance) of the distribution, is modeled as a multiplicative function of the covariates, and includes item ($\mu_i$) and store ($\eta_s$) fixed effects:

$$\ln \lambda_{ist} = \beta^{NB} (\text{discount}_{is} \times \text{national brand}_i) + \beta^{PL} (\text{discount}_{is} \times \text{private label}_i) + \mu_i + \eta_s \quad (2)$$

The effect of the discount was estimated to increase sales by 16% for national brands, but only 3% for private labels (see multivariate results in Table 4). The difference between these estimates is statistically significant. Adjusting for the change in average prices in each condition, the point estimates of the price elasticities were $-1.22$ for national brands and $-0.25$ for private labels.

Table 4: Poisson Quasi-Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Discount Effect</th>
<th>Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Brand</td>
<td>0.156**</td>
</tr>
<tr>
<td>Private Label</td>
<td>0.031</td>
</tr>
<tr>
<td>Difference</td>
<td>$p &lt; 0.05$</td>
</tr>
</tbody>
</table>

* Significantly different from zero, $p < 0.05$

** Significantly different from zero, $p < 0.01$

Standard errors in parentheses, clustered on item.

Elasticities are the QML estimates scaled to adjust for change in prices due to the experiment.

3.3 Discussion

The field test reveals a substantial difference in price sensitivity, and the estimates for private label demand are surprisingly inelastic. A price elasticity between $-1$ and $0$ implies that by raising its prices, the retailer could generate higher revenues while selling fewer units, lowering total costs and increasing profit. A reason for not doing so may be to maintain the retailer’s price image, due to concerns about store traffic. Furthermore, these estimates measure price sensitivity at price levels in the neighborhood of the observed regular prices, and may not predict how price sensitivity would change if prices were increased by large amounts, or if many prices were increased at the same time.

Private label price sensitivity may increase dramatically as the price gap with national brand prices increases.

---

As a robustness check, we estimated instrumental variables models that explicitly control for compliance in pricing conditions, which resulted in slightly larger magnitudes but did not have a substantive impact. Also, while the dollar value discount was larger on average for national brands, we obtained similar results using price levels as the independent variable. We performed a falsification test by applying the model to an earlier 17-week period, and the difference between conditions was not significant. Several alternative specifications, including a negative binomial model and zero-inflated Poisson model, did not result in qualitative differences.
narrow, but the retailer has explicitly stated that they take care to maintain the gap.

The difference in price sensitivities favors predictions that price sensitivity should be lower for private labels than for national brands. These include selective entry by the retailer, a lack of substitutes for the cheapest product, and lower customer price knowledge for private labels. A limitation of this experiment is that it does not permit us to distinguish between these explanations. Furthermore, these factors, as well as the offsetting factors that predict higher price sensitivity for private labels, are not mutually exclusive; each may play a role to varying degrees for different products. An investigation into the relative importance of these factors is beyond the scope of the paper, but is the subject of ongoing research.

A separate question raised by this finding is why the estimates differ from previous research. The point estimates of $-1.22$ and $-0.25$ are relatively low in magnitude, though they fall within the ranges reported in the literature. In Bijmolt, Van Heerde, and Pieters (2005), 15% of reported elasticity estimates fall between 0 and $-1$, and 25% fall between $-1$ and $-2$; the average elasticity was $-2.62$. A likely explanation for why the estimates in this study are lower than average is that we measure the effect of changes in regular price. Previous studies have found that regular price elasticities are relatively low, whereas promotional elasticities are higher in the short run, and lower in the long run (Jedidi, Mela, and Gupta, 1999; Pauwels, Hanssens, and Siddarth, 2002). Time-shifting and stockpiling effects generate strong responses during promotions, and sales may drop below normal after promotions end (Hendel and Nevo, 2006). Other marketing actions, such as in-store displays, sale signs, and advertising, may be correlated with these temporary discounts; when these factors are difficult to disentangle, promotional price elasticity estimates may increase further. We will reexamine this proposed explanation when we discuss the results of the instrumental variables estimation.

Another feature of the experimental results is that the difference between private label and national brand price sensitivity is more pronounced than in previous research. Studies using broad samples of products, including the Bijmolt et al. (2005) meta-analysis, have not found significant differences. The endogeneity of prices could account for this null result, particularly if retailer pricing policy differs for private labels and national brands. For example, consider a case where the private label price is set to maintain a certain gap with the national brand price. A positive demand shock specific to the national brand could induce a price increase in both products, while also causing the private label to lose share to the national brand. This would overstate the price sensitivity of the private label. This endogeneity problem potentially affects many empirical estimates in the literature. The field experiment removes concerns about endogeneity of prices because prices are exogenously manipulated.
A drawback of the field experiment is the comparatively narrow sample, in the sense that 192 items is a small fraction of the thousands of items that this retailer sells. Yet running an experiment on this scale is already a considerable undertaking. Many retailers may be reluctant to run similar studies, due to the direct costs of implementation and the loss of profits from suboptimal pricing. Also, increasing the breadth of such an experiment could cause interference between the items included in the sample. Historical data generally has the advantage of sampling broadly from items across many stores, and is readily available to researchers and firms. However, there is a pressing need for reliable validation of non-experimental methods for estimating price effects in such data. By using the experimental result as a benchmark, we can evaluate the different identification strategies. Furthermore, extending the analysis to the historical data can help show how correcting for endogeneity accounts for the departure from previous findings.

4 Instrumental Variables

The remainder of this paper presents empirical estimates of private label and national brand price elasticities using the firm’s transaction history. In this section, we use instrumental variables (IV) estimation to correct for endogeneity of prices. First, we introduce the historical data and report linear regression estimates that do not correct for endogeneity. Then we will consider using wholesale prices and commodity prices as instruments, and discuss the merits and limitations of each. We conclude this section with comments on the results of the IV estimation.

4.1 Data

The transaction data was obtained at the SKU level. Sales are aggregated over 81 stores from the same retail chain as in the field experiment, with 348 weekly observations from 2003-2009. The information available includes regular prices, paid prices, costs (wholesale prices), and quantities sold for each week. We performed analysis on the balanced panel of 1,969 items which were available throughout the time horizon, for a total of 685,212 observations. This set included 493 private label items (25%).

Item-level data identifies competitive items for some private label items, and for national brands targeted by private label brands. A retailer-defined SKU hierarchy was also available, with information on product categories.
4.2 Model

We estimate price sensitivity by regressing log quantities $Q$ on prices $P$, using the equation below. The price coefficients can be interpreted as elasticities. The subscript $i$ indexes items, and $t$ indexes weeks. We use a linear regression, which will permit a straightforward specification for instrumental variables estimation (with aggregate data, the larger quantities and lack of observations with zero demand weaken the motivation for the Poisson regression). The model includes item ($\mu_i$) and week ($\delta_t$) fixed effects. The item effects control for differences in scale between items with high and low sales volume, and for any item-specific factors that do not vary over time. The week fixed effects control for factors that only vary over time, and account for trends in aggregate demand. Variations in the specification for time (such as using a time trend instead of dummies, or estimating a separate trend for private labels) do not qualitatively change the estimates.

\[
\ln Q_{it} = \beta^{NB} (\ln P_{it} \times \text{national brand}_i) + \beta^{PL} (\ln P_{it} \times \text{private label}_i) + \mu_i + \delta_t + \varepsilon_{it} \tag{3}
\]

The estimates pool across items, to compare average price sensitivity for all 1,969 products. Interactions between price and dummy variables (for private labels and national brands) are used to separately estimate the price sensitivity of private labels and national brands. This results in measures of average price elasticity that are directly comparable to the results of the field experiment. Pooling is also helpful because price changes for individual items occur relatively infrequently. For a given item, the regular price will tend to change once a year or less. Thus, despite the long time frame, we observe very little variation in prices, which makes it difficult to estimate price sensitivity for individual products.

4.3 OLS Estimates

The equation above was first estimated using ordinary least squares (OLS). The point estimates of the regular price elasticities were $-1.08$ for national brands, and $-0.88$ for private labels (see Table 5). The estimates are significantly different from zero, but the difference between these coefficients is not significant. This pattern does not replicate the results of the field experiment. A possible reason for this inconsistency is that the OLS estimates are vulnerable to bias as a result of price endogeneity. As previously noted, endogeneity could account for the mixed results in previous studies, and our OLS estimates are similarly ambiguous. To reconcile the historical data with the

---

6The reported standard errors are clustered on item and week. Clustering on item provides results that are robust to arbitrary correlation in the error term for different time periods, while clustering on week provides results that are robust to correlation in errors for different products. This adjustment is used because a product’s weekly sales could be serially correlated, and contemporaneous errors could be correlated for related products.
Table 5: OLS Estimates of Price Elasticity

<table>
<thead>
<tr>
<th></th>
<th>Experimental Benchmark</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>−1.22**</td>
<td>−1.08**</td>
</tr>
<tr>
<td>Brand</td>
<td>(0.38)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Private</td>
<td>−0.25</td>
<td>−0.88**</td>
</tr>
<tr>
<td>Label</td>
<td>(0.25)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Difference</td>
<td>* p &lt; 0.05</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

* Significantly different from zero, p < 0.05
** Significantly different from zero, p < 0.01
Standard errors in parentheses, clustered on item and week.

field experiment, we next turn to instrumental variables estimation.

4.4 Wholesale Price as an Instrument

Instrumental variables estimation is a powerful tool for causal inference. This method uses a variable known as an instrument, which affects the dependent variable (quantity) only through the endogenous regressor (price). A valid instrument should be correlated with the endogenous regressor, and uncorrelated with the statistical error term (commonly referred to as the exclusion restriction). If the assumptions hold, IV allows us to use variation in prices correlated with cost changes to identify the effect of prices on demand, while ignoring variation in prices correlated with omitted factors. IV is frequently used in research requiring consistent estimates of price parameters.

The wholesale price provided by the retailer is an attractive potential instrument. The retailer prioritizes the proper computation and maintenance of this variable, updating it to reflect the firm’s current replacement costs as provided by their supplier contracts. Thus, it is a good proxy for marginal cost, the economic construct that helps determine optimal prices. In practical terms, the firm’s managers rely on the cost variable to compute margins for each item, and as the basis for making pricing decisions. The variable does not include trade promotions or other short-term changes in cost, so it should not cause short-term changes in price due to promotions. Thus, we expect that changes in price due to changes in this measure of cost are comparable to the unadvertised regular price change implemented in the field experiment. Inspection of the price and cost time series for individual items shows that price changes are often precipitated by cost changes, and the two variables are highly correlated. First-stage statistics are reported in Table 6.

The second assumption, the exclusion restriction, is difficult to assess. One concern is that wholesale prices may respond to a number of unobserved factors affecting demand, such as man-
Table 6: First-Stage Estimates (Regression of Retail Prices on Costs)

<table>
<thead>
<tr>
<th></th>
<th>Wholesale Price</th>
<th>Proc. Food PPI</th>
<th>Pharma. PPI</th>
<th>Angrist-Pischke F</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>0.77**</td>
<td>0.12**</td>
<td>0.57**</td>
<td>179.2</td>
</tr>
<tr>
<td>Brand</td>
<td>(0.06)</td>
<td>(0.02)</td>
<td>(0.07)</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>Private</td>
<td>0.22**</td>
<td>0.18**</td>
<td>0.72**</td>
<td>37.2</td>
</tr>
<tr>
<td>Label</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.11)</td>
<td>$p &lt; 0.01$</td>
</tr>
</tbody>
</table>

* Significantly different from zero, $p < 0.05$

** Significantly different from zero, $p < 0.01$

Standard errors in parentheses, clustered on item and week.

First-stage estimates based on model using full set of instruments.

Results are qualitatively similar for models using wholesale price and commodity prices separately.

manufacturer advertising expenditures. This could result in a simultaneity problem. Another concern is that wholesale prices affect demand through mechanisms other than price. For example, if the retailer adjusts the prices of competitive products in reaction to cost changes, it could affect demand through cross-price effects. Wholesale price could also affect other, harder to observe actions, such as shelf space allocation. Prior research on consumer goods has assumed that changes in wholesale prices are mainly driven by manufacturing costs. Chintagunta (2002) argues that manufacturers are required by law to set market-level wholesale prices, and so cannot easily react to changes in demand specific to a retailer. However, given the plausible explanations for how exclusion could fail, any argument in favor of using wholesale prices rests on the expectation that the correlation with retail prices is very high relative to potential violations of this assumption.

We present several approaches to evaluate the exclusion restriction. After introducing additional instruments later in this section, we report specification tests that favor the validity of the wholesale price as an instrument (see Table 7). However, we recognize that these tests are of limited diagnostic value.

An alternative approach is to perform a sensitivity analysis, in which we relax the exclusion restriction (e.g. Angrist, Imbens, and Rubin, 1996). Our analysis shows that the estimates are robust to small amounts of correlation between the wholesale price and the statistical error (see Appendix). Finally, we will use the field experiment as a benchmark to evaluate the IV estimates.

The model was re-estimated using wholesale price as an instrument for retail price. The estimated price elasticities were $-1.19$ for national brands, and $-0.16$ for private labels (see Table 7).

It has been suggested that an over-identified model, with more instruments than endogenous variables, permits a degree of specification testing of the exclusion restrictions (Wooldridge, 2002, §6.2.2). Hansen’s $J$ statistic is a general specification test for GMM estimators, and we fail to reject the null of no misspecification. We caution that this procedure cannot reliably refute or validate exclusion restrictions; the test’s power is unpredictable, and different instruments may identify different treatment effects.
The IV estimates are nearly identical to the estimates from the field experiment, and private labels are significantly less price sensitive than national brands. This strongly suggests that, despite concerns about the exclusion restriction, the wholesale price instrument eliminates far more bias than it introduces, and appears to be an effective way to estimate price sensitivity.

### 4.5 Commodity Prices as Instruments

Commodity prices that measure costs of manufacturing inputs have also been used as instruments. Market-level indexes are unlikely to have a direct relationship with demand in our data, and thus it is more likely to satisfy the exclusion restriction. However, they are less correlated with retail prices. We evaluated several producer price index (PPI) series as potential instruments. PPIs for processed foods and pharmaceutical chemicals were used due to strong first-stage results and their relevance to the products represented at the store. The instruments each generated positive and significant first-stage coefficients, consistent with the intuition that higher costs result in higher retail prices, but the association is not as strong as it is for wholesale price (see Table 6).

The model was re-estimated using two-step GMM, with the commodity prices as instruments and also with a full set of instruments. With only the commodity prices as instruments, the estimated price elasticities were $-1.89$ for national brands, and $-0.20$ for private labels. The difference between these coefficients is statistically significant, which replicates the corresponding result from field experiment. However, the estimates are less precise and differ in magnitude from the experimental benchmark; the point estimate for national brand elasticity differs from the experimental benchmark by two standard errors (see Table 7).
4.6 Discussion

While the OLS estimates were inconsistent with the field experiment, the IV estimates appear to do well in reproducing the experimental results. Using the wholesale price as an instrument yields very similar findings to the field experiment. Results using the commodity prices are credible but less precise. Notice that without the benefit of the experimental benchmark, there would have been little indication of which approach is more accurate. In many applications, there is a tradeoff between the strength of correlation with the endogenous regressor and confidence in the exclusion restriction. These two factors affect the precision and bias of the IV estimator. In this case, the strength of correlation favors the wholesale price instrument, while the confidence in the exclusion restriction favors the commodity prices.

A comparison of the IV and OLS results indicates that the difference in magnitude between our elasticity estimates and those typically reported in the literature can at least partly be attributed to our focus on regular prices. The historical data contains paid prices, which include temporary price promotions. For some items, the paid prices can vary substantially from week to week, and are correlated with short-term fluctuations in quantities, whereas for other items it deviates very little from the regular price. Paid price elasticities estimated using OLS are much higher than regular price elasticities (−2.10 for national brands and −1.64 for private labels), and correspond more closely to the elasticities in other studies. However, when using the wholesale price as an instrument, the regular and paid prices yield identical elasticity estimates. The wholesale price does not appear to capture variation in prices due to promotional activity, and so provides estimates comparable to the field experiment results.

The IV estimates for the private label elasticity are lower in magnitude than the OLS estimate, whereas we typically expect IV estimates to be larger. This is because we expect firms to raise prices in response to demand increases, resulting in same-sign shifts in prices and quantities. This has the effect of attenuating demand elasticity estimates when not correcting for endogeneity. Our results, on the other hand, suggest that the endogeneity bias has the opposite sign for private labels. This effect could be a result of private label pricing policy. In some cases, the private label price may be set to maintain a certain gap with the national brand price. Given such a policy, a positive demand shock for the national brand could result in a price increase for both products, while also causing the private label to lose share to the national brand. As a result, OLS estimates would understate national brand price sensitivity, but overstate private label price sensitivity. There is some evidence that this could be occurring. An inspection of the price series for private label and national brand pairs shows that the prices often move together. The average correlation coefficient between the price series of these pairs is 0.60, while the average correlation between prices of products within
the same category (excluding paired products) is significantly lower, at 0.44. These pricing patterns also help explain the weaker first-stage relationship for private labels (see Table 6).

An alternative explanation for why the IV results differ for private labels and national brands is that the cost variables measure different local average treatment effects. This could occur if the retailer’s policy for passing on cost increases differs for private labels and national brands. If the retailer passed on private label cost increases only if customers are not price sensitive, while passing on cost increases for national brands regardless of demand conditions, the IV estimates for private labels and national brands may not be directly comparable. Such a policy could also help explain the weaker first-stage for private labels. However, this explanation would not affect the experimental estimates we use as a benchmark, and other likely explanations for the differences in first-stage coefficients would not affect the IV estimates.

One potential difference between the field experiment and the IV estimates is that the experiment was designed to avoid competitive factors. In the historical data, price changes for brands within a category could be correlated at both wholesale and retail levels. Since the historical data offers a larger sample of items than the field experiment, we were able to limit the analysis to paired data, consisting of national brands and private labels in direct competition (as defined by the SKU hierarchy). This procedure resulted in 398 private label and national brand pairs, which we used to evaluate how much accounting for cross-price elasticities would affect the own-price elasticity estimates. The own- and cross-price elasticities were estimated using three-stage least squares (3SLS). The own-price elasticities were \(-1.47\) for national brands, and \(-0.46\) for private labels; the estimates are significantly different from each other, but only the former was significantly different from zero. Of the two cross-price elasticities, only the effect of national brand prices on private labels was significantly different from zero, at 0.58. It was not significantly different from the effect of private label prices on national brands, at 0.29. Although we have not specified a full system of cross-elasticities (for which it can be difficult to generate reliable estimates), these paired estimates

---

8Under conditions established by Imbens and Angrist (1994), the IV estimator can be interpreted as a local average treatment effect (LATE). In particular, the instruments should satisfy a ‘monotonicity’ assumption; in this case, we expect higher costs to result in weakly higher prices. A LATE measures the average demand elasticity, weighted towards units of observation whose prices are more responsive to changes in cost (referred to as ‘compliers’). For example, consider a retailer selling two products. If the retailer passes through 100% of costs for the first product, and 0% for the second product, IV would measure a LATE that captures the effect of price changes for only the first product. In this manner, differences in the set of compliers could drive the difference between private label and national brand IV estimates.

9In particular, a percentage change in wholesale price is smaller in absolute terms for a private label product than for a national brand product. There are fixed costs associated with changing prices, and they do not necessarily scale with price levels. These include direct menu costs (Levy, Bergen, Dutta, and Venable, 1997, and Anderson, Jaimovich, and Simester, 2011) and the risk of antagonizing customers (Anderson and Simester, 2010). A larger percentage change in wholesale price may be necessary before it is worth incurring these costs for private labels as compared to national brands. As a result, private label retail prices would be less responsive to wholesale prices, accounting for the differences in first-stage coefficients.

10These estimates are consistent with the asymmetry previously observed for promotional prices (e.g. Blattberg and Wisniewski, 1989), but do not provide strong evidence that this effect extends to regular prices.
suggest that the agreement between the experimental and IV estimates would not be disrupted by accounting for competitive effects.

Another approach to comparing estimates using the historical data to the field experiment would be to treat price changes occurring in the historical data as quasi-experiments. The impact of the price changes could be evaluated for short time frames, to avoid the effects of changing demand over time and competitive effects. Using this rationale, we next turn to a regression discontinuity design.

5 Regression Discontinuity

In this section, we describe a procedure for estimating price elasticities using a regression discontinuity design. We discuss identifying assumptions, and then use structural break methods to locate price change events that might satisfy these assumptions. We then report the regression discontinuity estimates, and compare them to the previous results.

Regression discontinuity (RD) identifies a treatment effect based on a discontinuity in the probability of treatment as a function of a running variable (also referred to as the forcing variable or scoring variable). The estimates are produced by comparing treatment and response on either side of the discontinuity. For example, Hartmann, Nair, and Narayanan (2010) estimate the effect of targeted marketing based on the observation that firms apply marketing communication policies to discrete groups of customers. The underlying customer types, which are firms’ estimates of customer value, are continuous variables. As a result, customers near the threshold between groups are effectively randomly assigned to different marketing policies. Busse, Simester, and Zettelmeyer (2009) estimate the effect of employee discount pricing promotions on prices paid for automobiles. If the start date of the promotion is not confounded by sudden changes in demand conditions, transactions occurring near the start of the promotion are essentially randomly assigned to occur either before or during the promotion.

In this study, the treatment is price, and the running variable is time. The retailer’s prices are characterized by discrete changes, usually occurring once a year or less, so that they resemble a step function over time. Even if these price changes are correlated with unobserved demand factors, they are infrequent enough that we might expect the unobserved factors to be essentially constant during a short period near the price changes.

A valid RD design also requires that customers are unaware of price changes when making store visit decisions, or are uncertain about timing of price changes. Knowledge about the timing could cause some customers, who would have purchased before a price drop, to wait for the price to change before visiting the store (Hartmann, Nair, and Narayanan, 2010). In this study, these conditions
seem likely to hold. Regular price changes tend to be increases, which the retailer is unlikely to publicize through circulars or advertisements. Also, the retail format is one where many customers visit out of convenience or as a result of an acute need. These circumstances favor the assumption that the customers are not making sophisticated, well-informed store visit decisions that would invalidate the RD design.

5.1 Identifying Price Changes

The dataset does not include explicit price change dates, so these were inferred through analysis of the prices themselves. Price change events were identified by applying structural break detection methods to a linear regression of regular prices on dummy variables for each price regime. Bai and Perron (1998, 2003) provide a framework for testing for multiple structural breaks in linear models, and a dynamic programming algorithm to quickly estimate the set of dates that minimize the sum of square residuals (the alternative being a time-consuming grid search on all possible combinations of breakpoints).\footnote{We implemented this analysis using the ‘strucchange’ R package (Zeileis, Kleiber, Krämer, and Hornik, 2003).} The number of breakpoints for each price series is chosen to minimize BIC, a criterion which penalizes over-fitting.

Uncertainty in the estimated locations of the breakpoints indicates price change events that may be unsuitable for RD estimation. This is because we want to compare prices and quantities precisely at the point where the price change occurs. Thus, breakpoints with 95% confidence intervals any wider than the minimum interval (two weeks, or one week in each direction) were discarded (see Figure 1 for an example; the price change events on either side of week 150 were discarded). Also, to avoid confounding competitive effects, price changes that occurred within eight weeks of a breakpoint

Figure 1: Breakpoints for Necco Wafers
identified for a competitive product were also discarded. This requirement is consistent with the design of the field experiment, and had little effect on the results.

The final sample includes 3,410 price change events, occurring for 1,583 unique items. Of these, 385 were private labels, a proportion consistent with the original dataset, though private labels changed prices slightly less often than national brands. Most of the events (85%) were price increases. The majority of items had relatively few price change events: 86% had three or fewer over nearly seven years. The greatest number of events identified for a given item was eight, occurring for just two items.

5.2 Model

The data used for each event includes the 6 weeks of sales before and after the price change. The average prices and quantities are plotted against the running variable (the time since the price change, with time reversed for price decreases) in Figure 2. It is clear that average prices increase dramatically at the breakpoint, and quantities decrease. It also appears that there is movement in the average prices and quantities in the period preceding the estimated breakpoint, and that the levels have not fully adjusted in the period of the price change. This is most likely because the exact timing of the price changes do not coincide exactly with the discrete periods recorded in the data.

The imperfect timing of the price changes results in what is known as a “fuzzy RD” design, in contrast to a “sharp RD” design in which the treatment changes with certainty (see Imbens and Lemieux, 2008). To correct for this, we used dummy variables for the price change events as instruments for prices in estimating the equation below. Fixed effects are included for each price
change event, indexed by the subscript $j$, and for each week $t$.

$$
\ln Q_{jt} = \beta^{NB} (\ln P_{jt} \times \text{national brand}_j) + \beta^{PL} (\ln P_{jt} \times \text{private label}_j) + \nu_j + \delta_t + \varepsilon_{jt}
$$

(4)

5.3 Results

The results are reported in Table 8. Using a 6-week window and the basic specification, the estimated price elasticities were $-1.80$ for national brands, and $-1.46$ for private labels. As with the instrumental variables estimates, private label demand is significantly less price sensitive than national brand demand. However, the magnitudes of the elasticity estimates are higher under the regression discontinuity approach.$^{12}$

<table>
<thead>
<tr>
<th>Table 8: RD Estimates, 6-Week Time Window</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Experimental Benchmark</strong></td>
</tr>
<tr>
<td><strong>Price Changes Included</strong></td>
</tr>
<tr>
<td><strong>All Events</strong></td>
</tr>
<tr>
<td><strong>Cost Change, No Promotions</strong></td>
</tr>
<tr>
<td>National Brand</td>
</tr>
<tr>
<td>$-1.22^{**}$</td>
</tr>
<tr>
<td>(0.38)</td>
</tr>
<tr>
<td>Private Label</td>
</tr>
<tr>
<td>$-0.25$</td>
</tr>
<tr>
<td>(0.25)</td>
</tr>
<tr>
<td>Difference</td>
</tr>
<tr>
<td>$p &lt; 0.05$</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>39,168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cost Change, No Promotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1.80^{**}$</td>
</tr>
<tr>
<td>(0.07)</td>
</tr>
<tr>
<td>$-0.88^{**}$</td>
</tr>
<tr>
<td>(0.12)</td>
</tr>
<tr>
<td>$-1.46^{**}$</td>
</tr>
<tr>
<td>(0.09)</td>
</tr>
<tr>
<td>$-0.29$</td>
</tr>
<tr>
<td>(0.29)</td>
</tr>
<tr>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>$p &lt; 0.10$</td>
</tr>
<tr>
<td>40,920</td>
</tr>
<tr>
<td>7,464</td>
</tr>
</tbody>
</table>

* Significantly different from zero, $p < 0.05$
** Significantly different from zero, $p < 0.01$
Standard errors in parentheses, clustered on event.
The “Cost Change, No Promotions” subset is defined in Table 9.

There appear to be two reasons why the RD results differ. First, the short time window is sensitive to the presence of promotions. In particular, for a subset of private label products, the retailer increases promotions in advance of price increases, possibly to increase trial by customers before the regular price increases. This may mitigate the impact of the price increase in the long run, but by inducing forward-buying inflates short-run sensitivity to the price change. Secondly, the estimates are sensitive to the timing of price changes. For example, the retailer may choose to keep prices low during periods of seasonally high demand (Chevalier, Kashyap, and Rossi, 2003),

$^{12}$As an alternative specification, linear trends were fit for the period before and after the price changes. This is equivalent to fitting a local linear regression for price and quantity near the discontinuity point, and estimating price elasticity based on predicted prices and quantities on either side of the discontinuity point. However, the slight uncertainty in the timing of the price change favors estimation using only the means. To minimize the potential impact of inconsistent timing, the analysis was also run with the week on either side of the price change removed. Finally, the model was fit with time windows of varying widths. The alternative specifications did not qualitatively change the estimates.
and wait to apply a price increase as demand returns to baseline levels. As a result, even short estimation windows may pick up unobserved differences in demand. In contrast, the IV estimates remove the effects of promotions, and focus on variation in prices attributable to cost changes.

Table 9: RD Estimates, 6-Week Time Window, by Price Changes Included

<table>
<thead>
<tr>
<th>Promotional Activity</th>
<th>Matched to No Matching</th>
<th>Cost Change Cost Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Private</td>
<td>−0.88∗∗</td>
<td>−0.84∗∗</td>
</tr>
<tr>
<td>Brand</td>
<td>(0.12)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Private</td>
<td>−0.29</td>
<td>−1.17∗∗</td>
</tr>
<tr>
<td>Label</td>
<td>(0.29)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Difference</td>
<td>p &lt; 0.10</td>
<td>n.s.</td>
</tr>
<tr>
<td>N</td>
<td>7,464</td>
<td>10,440</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Promotional Activity</th>
<th>Matched to No Matching</th>
<th>Cost Change Cost Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Private</td>
<td>−2.35∗∗</td>
<td>−2.06∗∗</td>
</tr>
<tr>
<td>Brand</td>
<td>(0.15)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Private</td>
<td>−2.01∗∗</td>
<td>−1.54∗∗</td>
</tr>
<tr>
<td>Label</td>
<td>(0.20)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Difference</td>
<td>n.s.</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>N</td>
<td>9,276</td>
<td>13,740</td>
</tr>
</tbody>
</table>

* Significantly different from zero, p < 0.05
** Significantly different from zero, p < 0.01

Standard errors in parentheses, clustered on event.

Events were coded as promotional activity were those where paid prices (net of temporary price promotions) averaged at least 1% less than regular prices. Events were coded as matched to cost changes when a corresponding change in wholesale price occurred during the period from 6 weeks before until the week after the price change.

Table 9 re-estimates the model on subsets of the data, categorizing the price change events based on promotional activity and changes in wholesale price. On one dimension (the rows of Table 9), the sample was split based on whether the event was influenced by promotions during the estimation window; for events in the bottom row, average paid prices deviated by more than 1% from regular prices. On the other dimension (the columns of Table 9), the sample was split based on whether the price change occurred in proximity to cost changes. For events in the left column, the items had a same-sign change in wholesale price occurring during a window from six weeks before up until the week after the event; the hope is that the timing of these price changes is less likely to be determined by demand conditions. The estimates in the upper-left quadrant of Table 9 are closer to the experimental benchmark and IV estimates, and are slightly smaller in magnitude.
5.4 Discussion

The RD analysis treats discrete price change events as quasi-experiments in the short time frame near the events. As with the IV estimates, these estimates are local to the price levels when the change occurs, but, unlike the IV estimates, can also be influenced by the timing of the price changes and short run tactics employed by the retailer. These additional factors appear to have enough of an effect on demand during the estimation window that the RD design overestimates price sensitivity, as compared with the field experiment.

While it is the case that using a subset of the price changes can produce RD estimates consistent with the experimental benchmark, this is an adjustment made using cost data that was incorporated more efficiently in the IV analysis. By screening out much of the data, these estimates are inefficient and potentially introduce unanticipated biases. Finally, if we had only been able to compare the full sample RD estimates to the OLS estimates (Table 5), we might have been satisfied by realistic estimates built on reasonable-sounding assumptions, and adopted somewhat different estimates.

Despite these problems, the RD analysis complements the other estimation approaches. The time frames of the various estimates are slightly different: the field experiment provides a sustained 4-month comparison, with contemporaneous controls, whereas the RD analysis provides short-run snapshots. The IV estimates, in contrast, are based on the long run movements in costs, prices, and quantities over many years. Also, the decomposed RD results help illustrate how the IV analysis removes the effects of promotions and the timing of price changes.

One final concern is that the price effects in the two datasets are not directly comparable: the experiment measures the effect of a regular price decrease, whereas the non-experimental findings are primarily based on price increases. We might expect these elasticities to vary due to asymmetric reference price effects (see Kalyanaram and Winer, 1995, for a review; Mazumdar, Raj, and Sinha, 2005, review evidence on competing explanations), which predict that the field experiment’s price decrease would result in attenuated price sensitivity estimates. Indeed, when the effect of the field experiment is estimated on a monthly basis, the national brand response does increase somewhat over the course of the 4-month period. However, the elasticities in the last month of the 4-month experiment (−1.58 for national brands and −0.17 for private labels) suggest that accounting for asymmetric reference prices would be unlikely to qualitatively change the experimental estimates. This is because reference prices are generally formed based on relatively recent observed prices (Briesch, Krishnamurthi, Mazumdar, and Raj, 1997, suggest 5-6 weeks), and should have adjusted by the last month of the experiment. The IV estimates are even less likely to be affected, due to the long-run nature of the prices in the historical data. Reference price effects could help account for the higher price sensitivities observed in the RD analysis, but the decomposition of the RD results
suggests that other factors played a larger role.

6 Conclusion

We have found that private label demand is less price sensitive than national brand demand. While previous findings have been ambiguous, we provide robust evidence from a U.S. retailer. The difference in price sensitivities between private labels and national brands may reflect a variety of factors, including market structure, heterogeneity in consumer preferences, and consumer perceptions of private label products. This finding suggests that lowering private label prices would not help grow private label demand, which is surprising given that market-level price sensitivity factors into long-run private label success (Lamey, Deleersnyder, Dekimpe, and Steenkamp, 2007). The main limitation of the results is that the estimates are local in nature, and may not predict the results of more drastic changes in pricing policy. In particular, the private labels in our study are nearly always priced lower than the competing national brand.

This study makes a methodological contribution by providing the first experimental validation of a conventional instrumental variables estimator. Previous empirical research has relied on wholesale prices and commodity prices as instruments. In particular, research using the Dominick’s Finer Foods dataset has used wholesale price as an instrument, although researchers have expressed concern about the exclusion restriction (e.g. Chintagunta, 2002). While the pros and cons of various instruments have been widely discussed, the validity of the available instruments cannot be determined using standard diagnostics. Our findings provide more confidence in the use of wholesale price as an instrument in similar retail settings.

Finally, this study belongs to a tradition of using field experiments to validate non-experimental methods. This paradigm has many potential applications in management research. The primary limiting factor is the availability of both experimental and non-experimental data that permit direct comparisons of estimation results. Overcoming the lack of data requires that academic researchers find partners in industry willing to run experiments. Researchers themselves must be willing to expose their own methods to more scrutiny. However, an agenda of validation, across a variety of settings, will provide the foundation for more robust non-experimental research.

13 The Dominick’s Finer Foods data is publicly available via the Kilts Center for Marketing, University of Chicago Booth School of Business.
Appendices

A Exclusion Restriction Sensitivity Analysis

To illustrate how the exclusion restriction might fail, and how this would affect the IV estimates, we performed a sensitivity analysis. This approach is described by Angrist, Imbens, and Rubin (1996). To provide a framework for this analysis, we relax the exclusion restriction, allowing cost \( C \) to affect quantity as in the equation below (we restrict our attention to wholesale price, since the exclusion restriction is more of a concern with this variable). The sensitivity parameters, \( \gamma \), represent the correlation between the cost variable and the error term when the exclusion restriction is imposed (Equation 3). The interpretation of \( \gamma \) depends on the posited source of endogeneity.

\[
\ln Q_{it} = \beta^{NB} (\ln P_{it} \times \text{national brand}_i) + \beta^{PL} (\ln P_{it} \times \text{private label}_i) \\
+ \gamma^{NB} (\ln C_{it} \times \text{national brand}_i) + \gamma^{PL} (\ln C_{it} \times \text{private label}_i) + \mu_i + \delta_t + \epsilon_{it} \tag{5}
\]

One potential source of endogeneity is simultaneity, where unobserved changes in demand affect wholesale prices. If positive demand shocks lead to higher costs, this would induce a positive bias in the cost coefficient in a reduced-form regression of quantity on cost. Since price and cost are positively correlated (as in the first stage regression), this results in a positive bias in IV estimates of \( \beta \). Thus, a simultaneity problem most likely would result in positive values of \( \gamma \). To the extent that national brand manufacturers have high bargaining power relative to private label manufacturers, this parameter is likely to be larger for national brands. To the extent that private label manufacturers sell a customized product to the retailer, and can more easily price discriminate, \( \gamma \) could be larger for private labels, but this scenario seems less likely.

Another possible source of endogeneity is if cost affects demand via a causal path other than a product’s own price. If higher costs cause the retailer to increase prices on substitute products, positive cross-elasticities would produce an upwards shift in demand, implying a positive value of \( \gamma \). Similarly, if higher costs cause the retailer to decrease prices on complementary products, negative cross-elasticities would also produce an upwards shift in demand.

Besanko, Dubé and Gupta (2005) found evidence of cross-brand wholesale price pass-through that was more likely to be positive for private label wholesale prices, implying \( \gamma \) is larger (more positive) for private labels. In our historical data, a regression of prices on competitive costs has a different pattern: national brand costs have a larger effect on private label prices (a coefficient of 0.36) than vice-versa (a coefficient of 0.05). Possible explanations for this difference are different retailer pricing policies, or the inclusion of trade promotions, though Dubé and Gupta (2008) did
not notice any systematic differences between regular price and promotional weeks.

With these explanations in mind for why the cost variable could violate the exclusion restriction, we estimate Equation 5 for several fixed values of $\gamma$. We benchmark $\gamma$ relative to the magnitude of the OLS pass-through estimates. The multipliers in the far right column are based on 3SLS estimates of cross-price elasticities. The corresponding adjusted price elasticity estimates are reported in Table 10.

**Table 10: IV Estimates Given Fixed Values of Gamma**

<table>
<thead>
<tr>
<th>$\gamma^{NB}$</th>
<th>$\gamma^{PL}$</th>
<th>National</th>
<th>Brand</th>
<th>Private</th>
<th>Label</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5 $\times$ PT</td>
<td>0 $\times$ PT</td>
<td>0.5 $\times$ PT</td>
<td>1 $\times$ PT</td>
<td>0.29 $\times$ PT</td>
<td>-0.5 $\times$ PT</td>
<td>0 $\times$ PT</td>
</tr>
<tr>
<td>-0.96**</td>
<td>-1.19**</td>
<td>-1.44**</td>
<td>-1.67**</td>
<td>-1.34**</td>
<td>(0.21)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>-0.04</td>
<td>-0.16</td>
<td>-0.31</td>
<td>-0.42</td>
<td>-0.33</td>
<td>(0.33)</td>
<td>(0.32)</td>
</tr>
</tbody>
</table>

* Significantly different from zero, $p < 0.05$
** Significantly different from zero, $p < 0.01$

Standard errors in parentheses, clustered on item and week.

PT is an OLS estimate of cross-brand pass-through of wholesale price.

The estimates for likely values of $\gamma$ provide a rough sense of the extent to which violations of the exclusion restriction might bias the results. We can see that intuitive explanations for such violations result in a minimal qualitative impact on the elasticity estimates. This analysis is reassuring, but does not allow us to conduct inference; each estimate assumes that $\gamma$ is known, and the standard errors are relatively unaffected.

As an extension of this sensitivity analysis, we used an informative prior distribution on $\gamma$ to construct confidence intervals accounting for uncertainty in the exclusion restriction. Conley, Hansen, and Rossi (2010) recommend this approach for situations where the instruments are ‘plausible.’ They offer wholesale price as an instrument for retail price as an example where such a procedure would be appropriate. In this section, we apply their method of constructing confidence intervals under the assumption that $\gamma$ (as parameterized in Equation 5) is near zero, rather than exactly zero.

A normal prior for $\gamma$ results in an analytical approximation for the asymptotic variance of the estimates, which assumes a known distribution for $\gamma$. The method is named for the "local-to-zero" (LTZ) assumption used to compute the approximation. Uncertainty about $\gamma$ is of the same order of magnitude as sampling error, so that neither source of error dominates the estimator’s asymptotic behavior.
Following Conley, Hansen, and Rossi (2010), Figure 3 shows price elasticity confidence intervals indexed by a parameter $\delta$, where $\gamma$ is distributed $N(0,\delta^2\beta^2)$, and Figure 4 shows confidence intervals for the difference between national brands and private labels. The components of $\gamma$ are assumed to be independent (a positive correlation would result in narrower confidence intervals). The difference between the parameters is significantly different from zero until two standard deviations of $\gamma$ span nearly 50% of $\beta$. A mean shift in the prior for $\gamma$ would result in a mean shift in the adjusted estimates for $\beta$, as in Table 10. The sensitivity analysis suggests that the difference in price sensitivity between private labels and national brands is robust to minor violations in the validity of the exclusion restriction.
References


