

Testing the Substitution Predictions of Demand Models: Evidence from the Blue Bell Recall

Wantian Huang*

University of Maryland

October 2026

Abstract

Demand models' substitution predictions inform merger review and the value of product variety. Using the 2015 withdrawal of the regional ice cream brand Blue Bell, I test how well models estimated on data through March 2015 predict the amount of demand retained by rivals and its allocation among them. Relative to markets Blue Bell never served, rivals recaptured about three quarters of its volume, and private label took half of that gain. Logit and random-coefficient models without nesting retain too little demand. A nested logit that pools private label with Blue Bell's mainstream tubs comes close both on how much demand stays with rivals and on which rivals receive it, and random-coefficient nested logit (RCNL), which allows consumer heterogeneity on the same nests, performs similarly. In this application, where private label sits in the nests matters more for these predictions than how much taste heterogeneity a model allows.

*Department of Economics, University of Maryland, College Park, MD 20742; whuang11@umd.edu.

1 Introduction

Estimates of substitution between products inform merger simulation and the valuation of product variety, including losses from recalls and stockouts. In *ex ante* applications that substitution comes from a demand model estimated before the change, and models that fit the same market shares can imply very different substitution. I evaluate these predictions using an observed whole-brand withdrawal: do standard demand models predict how much demand stays with rival products (retention) and which rivals receive it (allocation)?

In April 2015 Blue Bell Creameries recalled its entire product line after a listeriosis outbreak, and the brand disappeared from every retailer for several months. Blue Bell was a regional brand with about an eighth of ice cream volume in the markets where it sold. The withdrawal removed the whole brand at once, in every format and at every retailer, while its rivals stayed on the shelf, so the models' predicted substitution can be compared with the observed gains in rival sales.

I estimate logit, two nested logits that differ in the placement of private label, and three random-coefficient (RC) specifications on data through March 2015. The RC specifications, including random-coefficient nested logit (RCNL), use household micro moments to estimate the taste distributions (Petrin, 2002; Berry et al., 2004; Fan and Yang, 2025). The comparison thus varies both the grouping of products and the heterogeneity of tastes.

To compare the models' substitution predictions, I take the rivals' 2015 sales and prices from the data, return Blue Bell to each market at each model's forecast of its appeal, and measure the gallons its presence would have taken from each rival per gallon of its 2014 volume. Because rival prices are held at their observed values, the comparison needs no model of pricing conduct.

In Blue Bell's markets, rival products gained 0.77 gallons for each gallon Blue Bell had sold a year earlier, relative to markets where Blue Bell never sold, and private label took about half of that gain. Logit and the RC models without nesting keep less than half a gallon, while the nested logits and RCNL keep between 0.62 and 0.74. The nested logit that gives private label a group of its own sends it about a fifth of the gain, against half observed. The two nested specifications that pool private label with Blue Bell's mainstream tubs, nested logit and RCNL, are close on both margins: each has a misplaced share, a combined measure of errors in retention and allocation, of about a tenth of Blue Bell's volume, against about three tenths for the other four models. Their performance is similar: RCNL retains about as much as the nested logit and has a slightly smaller allocation error.

Among nested logits with nearly identical price coefficients, the grouping also moves welfare: at the rival prices observed after the recall, the one that gives private label its

own group values Blue Bell’s availability about 45 percent higher than the one that pools it with Blue Bell’s tubs. Rival prices add a term whose sign differs between observed and model-predicted prices: the models predict small increases in rival prices when Blue Bell leaves, whereas observed rival prices fell by 4.4 percent relative to markets without Blue Bell (Section 7).

This paper contributes to the literature on how the structure of a demand model shapes predicted substitution. Grigolon and Verboven (2014) estimate RCNL, which contains logit, nested logit, and random coefficients as special cases, and show that a model that omits a source of heterogeneity understates substitution within groups, although their merger predictions are fairly robust across the nested and random-coefficient specifications. Conlon and Mortimer (2021) show that diversion ratios from removals, price changes, and quality changes weight consumers differently and coincide only under plain logit or linear demand, and that diversion to the outside good is sensitive to omitted heterogeneity, in price or in the taste for inside goods, and to the market size. Conlon and Gortmaker (2025) show that, without variation in choice sets, first-choice demographic moments need not pin down unobserved taste dispersion or nesting parameters. I use the observed withdrawal to evaluate two sources of predictive error: how much of a removed brand’s demand stays with the remaining inside products, and how that demand is divided among them. The two nested logits differ substantially on both margins, whereas RCNL performs similarly to nested logit on the same PL-pooled nests.

It also contributes to work that evaluates demand models against observed changes in choice sets. Raval et al. (2022) estimate hospital demand on data before natural disasters and find that standard models underpredict large diversions after the resulting closures, Pathak and Shi (2021) compare demand-model forecasts of a school choice reform with realized outcomes, and Conlon and Mortimer (2013) use short stockouts in vending machines as variation that identifies substitution. Merger retrospectives compare simulated with realized price effects (Weinberg and Hosken, 2013; Björnerstedt and Verboven, 2016), so their errors combine demand and conduct. Bhattacharya et al. (2026) find, in 101 consumer-goods mergers, that predicted price changes are typically larger than realized ones, and when they re-invert post-merger demand at fixed parameters and the realized assortment, a check that goes back to Peters (2006), the merging parties’ price prediction error does not shrink. I add two things to this work. First, I distinguish the evaluation of substitution predictions at observed rival prices and sales from that of direct sales forecasts. Every model has a larger prediction error under the direct forecast, and the ordering of the non-nested specifications changes (Section 6.5). Second, I bring the test to a consumer-goods brand exit with posted prices, an outside option, and household purchase histories that follow the removed brand’s

own buyers through the withdrawal.

Work on product recalls shows why the observed rival gain needs care as a benchmark. Toy recalls lowered sales of non-recalled firms in the same category (Freedman et al., 2012), and product-harm crises can spill over to unaffected competitors (Cleeren et al., 2013), although rival brands showed no measurable sales losses after listeria recalls of branded ready-to-eat meats (Thomsen et al., 2006). A spillover of this kind would change rival sales through channels other than Blue Bell’s absence.

Sections 2 and 3 describe the market and measure the withdrawal. Sections 4 and 5 set out the models and their estimates, Section 6 the evaluation design and the comparison, Section 7 rival prices and consumer welfare, and Section 8 sensitivity. Section 9 concludes.

2 Setting and Data

2.1 The market and the recall

Ice cream reaches supermarket shoppers through a few national brands, many regional and small brands, and the retailers’ own private labels. Retailers’ private labels together hold 38.4 percent of 2014 volume across all DMAs. The national brands Breyers and Dreyer’s have 5.9 and 5.0 percent. The two premium pint brands, Ben & Jerry’s and Häagen-Dazs, have 5.1 percent together, and Blue Bunny has 2.3 percent. Other brands, many of them regional, hold 35.4 percent (Table 2).

Blue Bell was a regional brand. Blue Bell Creameries made ice cream, frozen yogurt, sherbet, and frozen snacks in several package formats, from three plants in Texas, Oklahoma, and Alabama.¹ The recalled products were sold at retail in 23 states (Velasco, 2015), most of them in the South and Southwest. In 2014 the estimation sample records Blue Bell sales in 105 DMAs, where the brand held 12.8 percent of ice cream volume.

The recall began with selected products on March 13, 2015 and extended to the entire product line on April 20 (Centers for Disease Control and Prevention, 2015). Production halted, and the withdrawal from stores was near complete from May through August (Figure 1, panel a). Blue Bell’s recorded gallons in its markets fell to 3 percent of their 2014 monthly mean in May and to less than 0.2 percent in June, July, and August; in the market-months the models are evaluated on, 0.7 percent of the 2014 May–August volume is still recorded, mostly in May. The return was staggered. Measured by the first month of

¹The products came in half gallons, quarts, pints, cups, and three-gallon containers and as take-home novelties (Wallace and Clayton, 2015), from plants in Brenham, Texas; Broken Arrow, Oklahoma; and Sylacauga, Alabama (Centers for Disease Control and Prevention, 2015; U.S. Food and Drug Administration, 2015).

renewed scanner sales, Blue Bell came back in 5 DMAs in September 2015 and in 7 more in October and November, then in 27 in December and in the rest during 2016, the last recorded re-entry among its DMAs falling in December 2016, and its recorded gallons stood at 40 to 80 percent of their 2014 monthly mean during 2016. The models are evaluated on the withdrawal; Appendix D describes household purchases after the return.

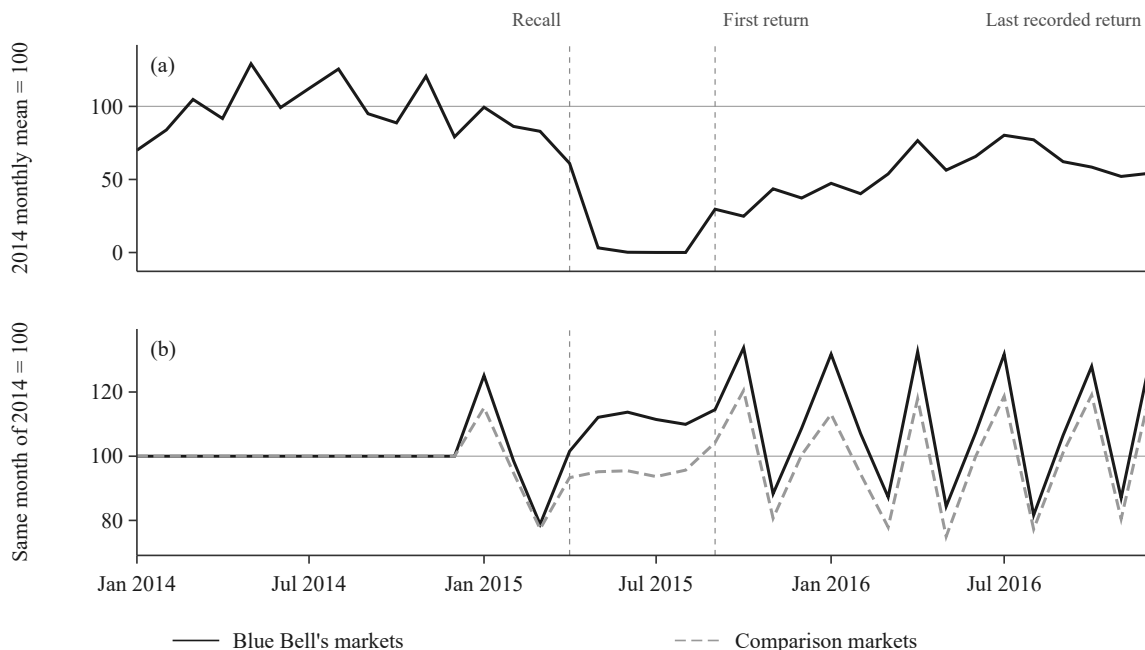


Figure 1: Blue Bell's gallons and rival gallons around the withdrawal

Scanner gallons in Blue Bell's markets and in the comparison markets, January 2014 to December 2016. Panel (a): Blue Bell's recorded gallons, 2014 monthly mean = 100. Panel (b): matched rival gallons, same month of 2014 = 100. Vertical lines: the complete recall (April 2015) and the first (September 2015) and last (December 2016) recorded returns among Blue Bell's DMAs.

A recall changes more than the choice set. News about the outbreak can shift demand for the category and for particular brands, and retailers can change shelf space, promotions, and private-label prices. The comparison takes three things as observed in 2015: rival prices, rival sales, and the sales of products outside the demand system. It does not remove the other changes the recall caused; Section 6.1 sets out how they enter the comparison.

2.2 Data, products, and the estimation sample

The data are the Nielsen retail scanner data for the ice cream, frozen yogurt, and frozen novelty modules, which record quantities and expenditures by product, store, and week, and the Nielsen Homescan panel, which records the purchases of the same households over time.

The unit of analysis is a product in a DMA–retailer market in a month. I use the scanner data to estimate market demand and to measure the response to the withdrawal.

A product is a brand combined with a set of characteristic indicators. Items that differ only in package size belong to the same product, and their quantities are converted into gallons, the common unit of quantity. The indicators describe the item, not its package size: bulk, for example, comes from Nielsen’s recorded form of the item rather than from a size threshold (Appendix A). Seven characteristics carry random tastes in the demand models (vanilla, chocolate, bulk, low fat, low calorie, assorted, and cookies or mix-ins).

Private-label products have no recorded values for the seven characteristics, and their brand names are masked.² A market is a designated market area (DMA) paired with a retailer, so the private-label product in a market is that retailer’s own store brand. Repeated observations of a product within a market and month are combined by summing expenditure and gallons, and price is expenditure per gallon.

The estimation sample contains 756 model products and 529 DMA–retailer markets in the 117 DMAs that pass the sample filters (Appendix A), from January 2011 through March 2015 (Table 1). March 2015 contains the first withdrawals, so the estimation window, data through March 2015, ends before the complete withdrawal but not before the event began.

Table 1: Estimation and evaluation samples

Sample	DMA–retailer markets (DMAs)	Observations
Estimation, January 2011–March 2015	529 (117)	3,705,958 product–market–months
Blue Bell markets, May–August 2015	317 (103)	1,246 market–months
Markets without Blue Bell in the counterfactuals	301 (71)	
Comparison markets for the observed effect	266 (65)	

Eligibility: estimation, model products with recorded sales; Blue Bell markets, Blue Bell scanner sales in May–August 2014; markets without Blue Bell, no Blue Bell scanner sales in May–August 2014 and no panel purchases (28 are in the estimation sample, and the others take market sizes by the rule of equation (A.1)); comparison markets, no Blue Bell scanner sales in any month of 2011–2015.

A market–month offers 137 model products on average (Table 2). The average product–market–month price is \$28.58 per gallon, while the volume-weighted average is only \$13.47, because volume is concentrated in low-priced family-size containers and private label, and many high-priced pints and novelties sell little.

I set the market size of each DMA–retailer to 1.5 times its largest monthly inside volume through March 2015 and hold it fixed over time (the full-window estimates use the months

²Nielsen masks store brands: retailers’ private-label items are recorded under four generic brand codes with no characteristic information, so each code is one private-label product with every characteristic indicator equal to zero (Appendix Table A.2).

Table 2: Summary statistics of the estimation sample

<i>Panel A. Coverage, January 2011 to March 2015</i>				
Model products	756			
DMA-retailer markets	529			
DMAs	117			
Months	51			
Product-market-months	3,705,958			
<i>Panel B. Prices and market-months</i>				
	Mean	10th pct.	Median	90th pct.
Price per gallon (\$), product-market-months	28.58	9.94	26.84	50.08
Price per gallon (\$), volume-weighted	13.47	7.12	9.97	28.29
Model products per market-month	137.4	49	128	233
Inside share of market size	0.403	0.264	0.399	0.556
<i>Panel C. 2014 volume shares by brand group (percent)</i>				
	All DMAs	Blue Bell DMAs		
Blue Bell	7.8	12.8		
Private label	38.4	41.2		
Breyers	5.9	5.8		
Dreyer's	5.0	4.1		
Premium (Ben & Jerry's, Häagen-Dazs)	5.1	4.3		
Blue Bunny	2.3	2.8		
Other brands	35.4	29.1		

Inside share is inside volume divided by the market size.

through 2016). The inside share is then at most two thirds in the largest month, and across all market-months it averages 0.40 (Table 2). The market size is an assumption about purchase opportunities, not a count of consumers, and it governs the split between the category and the outside option, so Section 8 repeats the comparison with a smaller market size. This choice is not an innocuous normalization: a misspecified market size distorts estimated preferences and substitution to the outside good, and random coefficients do not remove the distortion (Zhang, 2026).

The household panel serves two purposes. In estimation, its 2011–2014 purchases supply the micro moments that the RC models are fit to match (Petrin, 2002; Berry et al., 2004; Fan and Yang, 2025): how much buying households buy in a year, how this differs for buyers of products with each characteristic and across three income groups, and, for the group RC, how purchases concentrate within product groups (Section 4.3). In Section 3, it follows Blue Bell's own buyers from 2013 through 2016, in cohorts balanced on the panel roster, so that a change between two windows is not a change in which households report.

The evaluation covers May through August 2015, the months between the complete recall and the first return, with the same months of 2014 as the reference. The models are evaluated

on the 1,246 market-months in Blue Bell’s 317 DMA-retailer markets that had Blue Bell sales in the same month of 2014 and in which every counterfactual solved; all 317 markets are in the estimation sample. Table 1 maps the samples.

3 The Withdrawal in the Data

The scanner data provide the market-level benchmark against which Section 6 compares the models: a difference in differences in rival sales growth between Blue Bell’s markets and markets Blue Bell never served. The household panel follows Blue Bell’s own buyers through the withdrawal and provides complementary evidence on their purchases, from a trend-adjusted comparison rather than a direct measure of what the withdrawal did to them.

3.1 The scanner benchmark

The scanner benchmark counts gallons bought by all customers of covered retailers. For a group d of rival products and a market-month mt in which Blue Bell sold in the same month of 2014, the observed removal effect is

$$O_{dmt} = Q_{dmt,2015}^{obs} - r_d Q_{dmt,2014}^{obs}, \quad (1)$$

where both quantities sum over the product-markets that sold in that month of 2014, and r_d is the ratio of the group’s May–August 2015 gallons to its 2014 gallons in the DMA-retailer markets where Blue Bell never sold. The effect counts the rival gallons gained in Blue Bell’s markets beyond the change in the same products’ sales elsewhere. I sum it over the 1,246 market-months of Section 2 and divide by Blue Bell’s gallons in the same months of 2014. Intervals come from resampling DMAs, with their retailers and months, in Blue Bell’s markets and in the comparison markets together.

Rival products gained 0.77 gallons for every gallon Blue Bell had sold a year earlier (Table 3), with a 95 percent interval from 0.54 to 1.01. Private label took 0.40 of these gallons, about half of the gain, and Dreyer’s took 0.16. Other brands took 0.14, while Breyers, the two premium pint brands, and Blue Bunny each took about 0.03 or less. Dreyer’s share of the gain is about twice what allocation in proportion to rival sales in the same market-months would give it (0.21 against 0.11). Listings that were new in 2015 add another 0.13 gallons, not included in the 0.77, mostly in small brands outside the estimated product set.

Rival sales gains are present in every month of the withdrawal. Rival gains per 2014 Blue Bell gallon are 0.71, 0.92, 0.77, and 0.73 in May through August (Figure 2), and their mean weighted by Blue Bell’s 2014 volume is the 0.774 of Table 3.

Table 3: Observed rival gains in Blue Bell’s markets, May–August 2015

	Gallons per 2014 Blue Bell gallon	95 percent interval	Share of the gain
Private label	0.398	[0.176, 0.569]	0.515
Dreyer’s	0.160	[0.119, 0.246]	0.207
Other brands	0.137	[0.088, 0.238]	0.177
Breyers	0.033	[−0.026, 0.059]	0.042
Premium pints	0.031	[0.022, 0.038]	0.040
Blue Bunny	0.014	[−0.009, 0.033]	0.018
All rivals	0.774	[0.544, 1.006]	1.000
<i>Not included above</i>			
Listings new in 2015	0.128	[0.068, 0.217]	
Placebo, 2013 to 2014	0.123	[−0.019, 0.347]	

Premium pints are Ben & Jerry’s and Häagen-Dazs.

Figure 3 places the withdrawal in the longer scanner history. It plots the relative growth of rival sales in Blue Bell’s markets against the comparison markets, each month against the same month of 2014, from January 2013 through December 2015. Because each month is its own reference, 2014 is zero by construction, and the information before the event is in 2013 and in January to March 2015. Through most of 2013 the relative growth is a few percent below zero, so rivals in Blue Bell’s markets grew faster into 2014, which is what the placebo of Table 3 picks up. The relative growth is 9, 4, and 2 percent in January to March 2015, 18 to 19 percent in May to July and 15 percent in August, and about 10 percent in the months after the first returns. It rises sharply with the complete recall, though it was already positive in early 2015.

Exposure is Blue Bell’s share of DMA volume in 2014. The gain per Blue Bell gallon is substantial in every exposure tier and smaller in the high-exposure tier: 0.95 in the low tier, 0.91 in the medium tier, and 0.66 in the high tier (Table 4). The low tier’s interval, from 0.14 to 2.03, is wide, so the ordering is not established.

The effect measures what the recall did to these rivals’ sales under three conditions:

1. without the recall, each destination group’s sales in Blue Bell’s markets would have grown in the same proportion as in the comparison markets;
2. the recall did not change sales in the comparison markets;
3. the scanner data measure sales in the same way in both sets of markets.

The benchmark is also one event, so its interval describes dispersion across DMAs around the recall that occurred, not variation across recalls. The same construction applied to

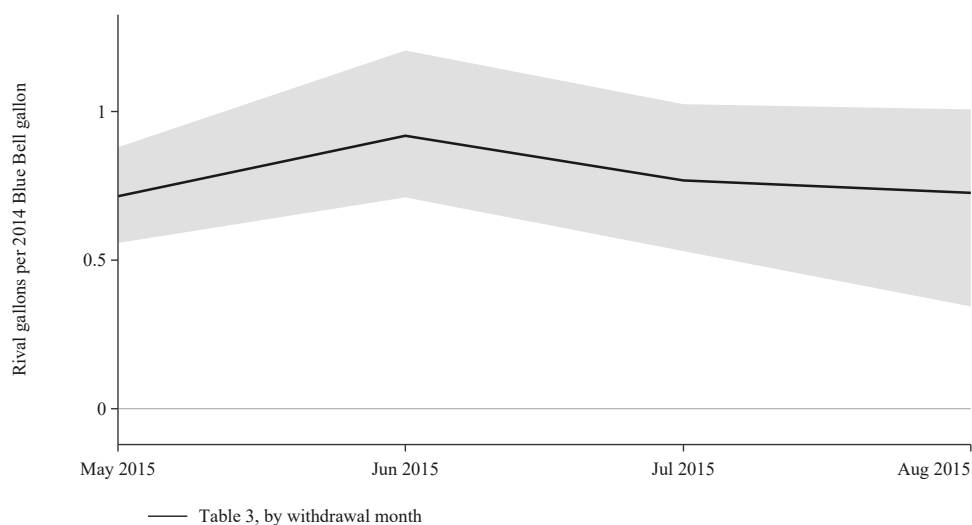


Figure 2: Rival gains per 2014 Blue Bell gallon by withdrawal month
Equation (1) by month, with the comparison markets' growth rates pooled over May–August as in Table 3; 311, 308, 313, and 314 of the 1,246 market–months. Shaded band: pointwise 95 percent intervals from 199 DMA block-bootstrap draws.

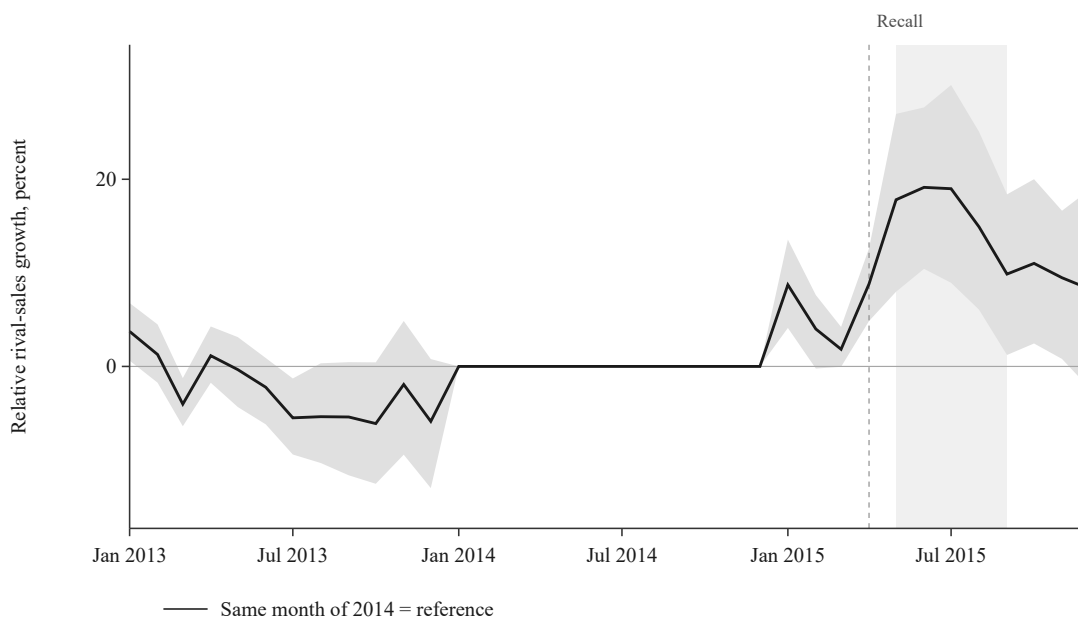


Figure 3: Relative growth of rival sales in Blue Bell's markets
Growth of matched rival gallons in Blue Bell's markets over the same month of 2014, relative to the same growth in the comparison markets, in percent. The May–August 2015 months, shaded, use the benchmark's 1,246 market–months; the other months extend the same matching rule to the scanner history, so their support differs. Band: pointwise 95 percent intervals from 199 DMA block-bootstrap draws.

Table 4: Observed rival gains by Blue Bell exposure

	Low	Medium	High
Private label	0.368 [−0.318, 1.237]	0.466 [0.229, 0.678]	0.371 [0.288, 0.420]
Dreyer’s	0.365 [0.209, 0.622]	0.176 [0.084, 0.268]	0.103 [0.060, 0.117]
Other brands	0.398 [0.153, 0.931]	0.124 [0.028, 0.265]	0.082 [0.033, 0.102]
Breyers	−0.171 [−0.591, −0.013]	0.079 [0.054, 0.106]	0.057 [0.041, 0.063]
Premium pints	0.025 [−0.006, 0.041]	0.032 [0.019, 0.046]	0.032 [0.026, 0.038]
Blue Bunny	−0.041 [−0.077, 0.002]	0.033 [−0.002, 0.078]	0.018 [0.012, 0.031]
All rivals	0.946 [0.144, 2.028]	0.910 [0.606, 1.216]	0.664 [0.494, 0.709]
DMAs	35	34	34
Market-months	355	471	420

Gallons gained per 2014 Blue Bell gallon in the 1,246 market-months of Table 3, by Blue Bell's 2014 share of DMA volume (cutoffs 7.78 and 31.27 percent), with the comparison markets' growth rates of Table 3, shared across tiers. Brackets: 95 percent intervals from 199 DMA block-bootstrap draws.

2013 and 2014, when nothing was removed, gives 0.12 per Blue Bell gallon, with an interval from -0.02 to 0.35 . Part of the 0.77 may therefore reflect rivals that were already growing faster in Blue Bell’s markets, a threat to the first condition, and Section 8 scores the models against the effect net of this placebo. The placebo checks comparability before the recall; it cannot rule out changes after the recall, such as a scare that moved the whole category, which Section 6.1 discusses.

3.2 Blue Bell’s buyers during the withdrawal

The household panel shows the same event from the side of Blue Bell’s own buyers. They replaced a little over half of their former Blue Bell volume with rival products. Net of the trend of comparable households, they bought 0.555 more gallons of rival products in the estimation universe for every gallon of Blue Bell they had bought in May–August 2014 (standard error 0.043; Table 5). So that households are not selected on the purchases that measure the volume removed, the cohort is the households that bought Blue Bell in 2013 and stayed on the panel roster through 2015; the trend comes from households in the same DMAs that bought no Blue Bell in 2013 or 2014.³

The households’ gains went disproportionately to mainstream branded tubs. Private label received 41.6 percent of the net rival increment, slightly less than its share of the same

³Selecting instead on Blue Bell purchases in May–August 2014 gives a lower net retention, 0.475 (standard error 0.020), from a cohort selected on unusually high baseline purchases (Appendix Table D.1). Counting all ice cream rather than the estimation products raises the estimate to 0.595, so about 0.04 gallons per Blue Bell gallon went to products outside the estimated product set, which a model of the estimation products counts as the outside option. The same construction applied to the year before the recall gives -0.017 .

Table 5: Net retention and allocation among Blue Bell’s buyers

	Estimate (SE)	2014 basket	Ratio
Net retention, estimation products	0.555 (0.043)		
Net retention, all ice cream	0.595 (0.046)		
Placebo: 2012 buyers, 2013 to 2014	−0.017		
<i>Share of the net rival increment, estimation products</i>			
Private label	0.416 (0.036)	0.460	0.9
Blue Bunny	0.161 (0.019)	0.084	1.9
Breyers	0.094 (0.016)	0.062	1.5
Dreyer’s	0.078 (0.015)	0.046	1.7
Ben & Jerry’s	−0.002 (0.006)	0.024	−0.1
Häagen-Dazs	0.017 (0.008)	0.017	1.0
Other brands	0.237 (0.028)	0.307	0.8

Standard errors in parentheses (Appendix D gives the bootstrap). The cohort has 5,429 households, restricted to those that bought ice cream in May–August of 2014 or 2015. The trend $\hat{g} = 0.964$ comes from 22,280 households. The 2014 basket is the cohort’s May–August 2014 rival purchases, and the ratio is the first column over the second. The placebo applies the same construction to the 5,257 households that bought Blue Bell in 2012, from May–August 2013 to May–August 2014. Across the trend definitions of Appendix Table D.2, net retention within the estimation products ranges from 0.49 to 0.61.

households’ 2014 rival purchases, and Dreyer’s received 1.7 times its share (Figure 4). The two premium pint brands together received almost none. These are shares of the cohort’s own increment, a household allocation rather than the scanner allocation of Section 3.1.

Blue Bell’s buyers were no more likely than comparable households to stop buying ice cream. Among the cohort households that bought ice cream in May–August 2014, 9.5 percent bought none in the same months of 2015, against 10.1 percent among comparable households reweighted to the cohort’s 2014 category volume. The volume that rival purchases did not replace therefore came mostly from households that kept buying ice cream but bought less: excess non-purchase accounts for 0.009 gallons (standard error 0.010) of the 0.445 gallons per Blue Bell gallon that rival purchases did not replace, about 2 percent (Appendix Table D.3).

The two benchmarks measure different objects. The scanner effect counts every rival gallon gained in Blue Bell’s markets per gallon Blue Bell sold there. The household estimate counts what Blue Bell’s own buyers bought, net of comparable households, per gallon of their own Blue Bell purchases, so its level, 0.555, is not an estimate of the same quantity as the scanner effect’s 0.77. The two agree on the qualitative facts: most of Blue Bell’s volume went to rival products rather than out of the category, private label took about its share of rival purchases or a little less, premium pints less than theirs, and Dreyer’s more than its share. The scanner effect, which covers all buyers at covered retailers, is the benchmark the models are scored against.

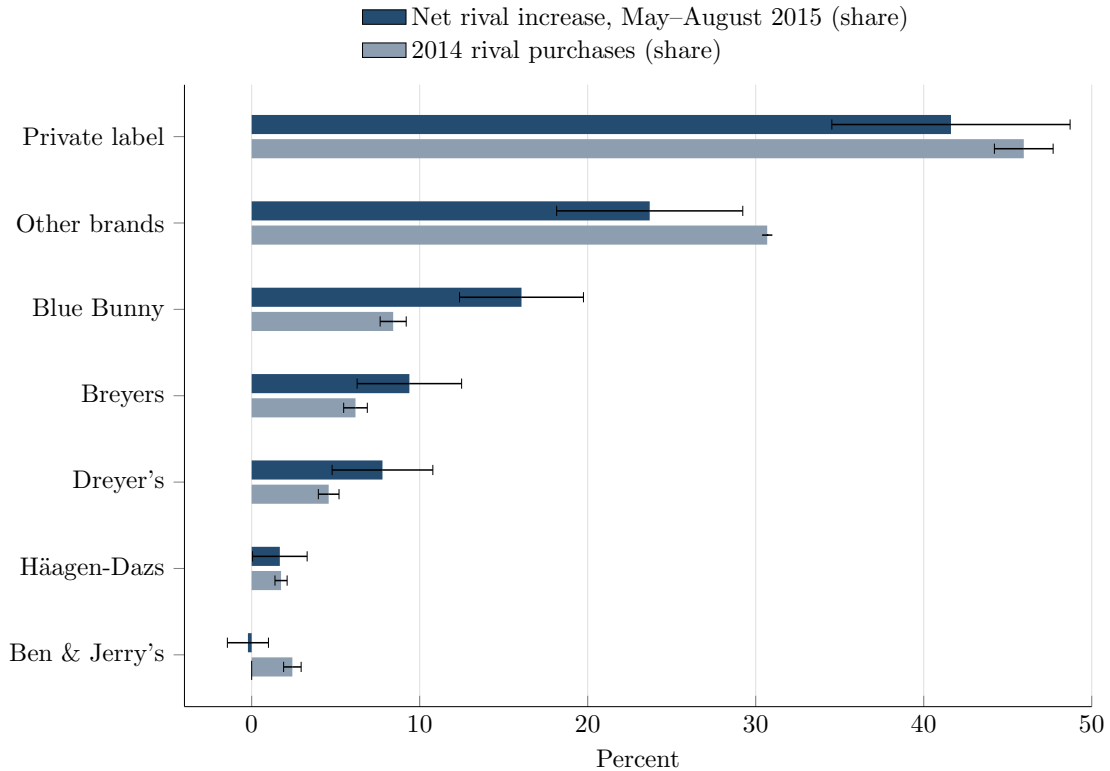


Figure 4: The net rival increment and 2014 rival purchases among Blue Bell buyers Cohort and products of Table 5; whiskers are 95 percent intervals, and the 2014 share of other brands has none.

Together, these facts require a model to capture both substantial retention and the uneven gains across rivals. Allocation in proportion to rival sales, the logit restriction, gets private label's share of the gain about right but gives Dreyer's about half of its observed share. Dreyer's gain also coincides with the largest relative price decline among the rivals (Section 7), so some of its gain may reflect the price cut rather than substitution at fixed prices.

4 Demand Models and Estimation

Every model in the comparison has the same economic setup. In each DMA–retailer market and month, a consumer buys one ice cream product or nothing. A product's appeal has a part common to all consumers (its mean utility, which absorbs price, product, market, and time effects) and, in the richer models, a part that varies across consumers. The models differ in how they represent substitution between products. Logit ties substitution among the

remaining products to their shares. Nested logit adds substitution within nests of products. Random coefficients let consumers who chose the same product differ in their preferences over its alternatives, and combining random coefficients with nesting allows both. I estimate every specification on the same products, markets, and months, and compare six of them on the withdrawal; Table 6 lists them and where they also differ in controls or micro moments.

4.1 Utility and substitution restrictions

Let j index products, m DMA–retailer markets, and t months. Consumer i chooses among the products available in the market and an outside option, indexed by zero. The utility of inside product j is $v_{ijmt} + \varepsilon_{ijmt}$ and that of the outside option is ε_{i0mt} , with

$$v_{ijmt} = \delta_{jmt} + \mu_{ijmt}, \quad (2)$$

$$\delta_{jmt} = a_j + c_m + \phi_{\text{year}(t)} + \tau_{\text{month}(t)} + \alpha p_{jmt} + \xi_{jmt}, \quad (3)$$

$$\begin{aligned} \mu_{ijmt} = & \sum_{k=1}^7 \sigma_k x_{jk} \nu_{ik} + \sigma_0 \nu_{i0} + \gamma_p l_i p_{jmt} \\ & + \lambda_{mid} \mathbf{1}[l_i \in \mathcal{L}_{mid}] + \lambda_{high} \mathbf{1}[l_i \in \mathcal{L}_{high}] + \eta_{i,g(j)}. \end{aligned} \quad (4)$$

Here a_j is a product fixed effect, c_m a DMA–retailer fixed effect, and ϕ and τ are year and month-of-year fixed effects. The RC model and the nested logits add retailer-by-private-label fixed effects to δ_{jmt} , which give each chain’s private label its own mean utility; logit, RCNL, and the group RC do not include these effects. They change the withdrawal scores little (Section 6.4). The price is p_{jmt} , in dollars per gallon, and ξ_{jmt} is the unobserved demand term, which varies within a product and market over time. The vector x_j contains the seven characteristic indicators: vanilla, chocolate, bulk, low fat, low calorie, assorted, and cookies or mix-ins. The tastes ν_{ik} and ν_{i0} are independent standard normal, so σ_k is the dispersion of the taste for characteristic k and σ_0 the dispersion of a common taste for all inside goods relative to the outside option. In logit and the non-nested RC specifications, the shocks ε_{ijmt} are independent across products and type-I extreme value conditional on the consumer’s type. In nested logit and RCNL, they follow the nested extreme-value structure of Grigolon and Verboven (2014):

$$\varepsilon_{ijmt} = \zeta_{ig(j)mt} + (1 - \rho) \tilde{\varepsilon}_{ijmt}, \quad (5)$$

where $\tilde{\varepsilon}_{ijmt}$ is independent type-I extreme value and the nest component ζ_{igmt} , common to the products of nest g , has the distribution that keeps ε_{ijmt} type-I extreme value (Cardell, 1997). The nesting parameter ρ thus governs the correlation of the shocks within nests, $1 - (1 - \rho)^2$, and the outside option forms a singleton nest.

Income enters through price sensitivity and through the level of utility for inside goods. The variable l_i is the log income of simulated consumer i , drawn as $l_i = \bar{l}_d + s_d z_i$ from the mean $\bar{l}_d$ and standard deviation s_d of log income in DMA d (Appendix B), with z_i standard normal. Consumer i 's price coefficient is $\alpha + \gamma_p l_i$, so α is its intercept and γ_p its change per unit of log income. The shifts λ_{mid} and λ_{high} move the utility of every inside good for consumers in the middle and high income groups relative to the low group; the sets $\mathcal{L}_{mid}$ and $\mathcal{L}_{high}$ use the same two log-income cutoffs as the household income moments. I use the two group shifts rather than a log income level because the household income moments are defined on these three groups: the shifts match the two income-quantity contrasts, while a log income level would add variation within groups that those moments do not see.

The specifications restrict equation (4) in different ways. Logit and nested logit set $\mu_{ijmt} = 0$. The RC model contains the seven characteristic tastes, the inside-good taste, γ_p , and the two income shifts. RCNL adds a common nesting parameter ρ , with the four PL-pooled groups defined below as its nests. The group RC adds a group-specific random coefficient $\eta_{ig} = \sigma_g z_{ig}$ for each of the same four groups, drawn independently across groups, and keeps σ_0 .

In logit and the non-nested RC specifications, the choice probability of a consumer of type i and the market share are

$$P_{ijmt} = \frac{\exp(v_{ijmt})}{1 + \sum_{k \in \mathcal{J}_{mt}} \exp(v_{ikmt})}, \quad s_{jmt} = \int P_{ijmt} dF_m(i), \quad (6)$$

where $\mathcal{J}_{mt}$ is the set of available products and F_m the distribution of types in market m . Without heterogeneity, the ratio of the probabilities of two remaining products does not change when a third product disappears. With heterogeneity, the ratio is unchanged within each type but can change in the market as a whole, because the removed product's buyers need not resemble the buyers of other products. A brand withdrawal provides a test of whether buyers move to remaining products in proportion to their shares or favor some of them.

For nested logit, let $\mathcal{J}_{gmt}$ denote the products in nest g and define the within-nest exponential sum

$$D_{igmt} = \sum_{k \in \mathcal{J}_{gmt}} \exp\left(\frac{v_{ikmt}}{1 - \rho}\right). \quad (7)$$

A consumer's probability of choosing product j is the probability of choosing its nest

Table 6: The six specifications compared on the withdrawal

Name	Substitution structure	Retailer ×	
		PL effects	Micro moments
<i>Panel A. Specifications</i>			
Logit	Proportional to shares	No	None
NL, PL separate	Common ρ within nests	Yes	None
NL, PL pooled	Common ρ within nests	Yes	None
RC	Characteristic and inside-good tastes	Yes	Base
RCNL	RC tastes plus common ρ , PL-pooled nests	No	Base
Group RC	RC tastes plus a taste for each PL-pooled group	No	Base and own-group
<i>Panel B. Groups of each grouping</i>			
PL separate	Mainstream tubs; premium tubs; mainstream novelties; premium novelties; private label		
PL pooled	Mainstream tubs and private label; premium tubs; mainstream novelties; premium novelties		
<i>Panel C. In every specification</i>			
Fixed effects	Product, DMA–retailer, year, and month of year		

NL is nested logit and PL private label. Micro moments: the base block has 12 moments and the own-group block 4 (Section 4.3). The excluded instruments are the characteristic sums, rival counts, and sugar-price columns of Section 4.3 in every specification except RCNL, which omits the sugar-price columns; the nested logits and RCNL add the within-nest instruments. Appendix Table B.1 lists the variants scored in the appendix.

times the probability of choosing j within that nest:

$$P_{ijmt} = \underbrace{\frac{\exp(v_{ijmt}/(1-\rho))}{D_{igmt}}}_{P_{ij|g,mt}} \underbrace{\frac{D_{igmt}^{1-\rho}}{1 + \sum_h D_{ihmt}^{1-\rho}}}_{P_{igmt}}. \quad (8)$$

The sum D_{igmt} is a computational device for the choice probability, not an additional utility parameter; $(1-\rho) \log D_{igmt}$ is the nest’s logsum, the expected maximum utility within the nest up to a constant, which Section 7 uses for welfare. The outside option is a singleton nest, and ρ , with $0 \leq \rho < 1$, is common to the inside nests. Nested logit and RCNL both use this conditional formula; RCNL integrates (8) over consumer types, whereas nested logit has no random tastes.

The groups express hypotheses about which products are close substitutes. I build two groupings from two attributes measured before the withdrawal: format, which separates novelties (bars, cones, and sandwiches) from packaged ice cream in cartons, tubs, and pints, which I call tubs; and price tier, under which a brand–format is premium if its pre-withdrawal price exceeds \$20 per gallon for tubs and \$25 for novelties.⁴ Both groupings split tubs and

⁴Among the tub brand–formats, no pre-withdrawal price lies between \$19.11 and \$20.64 per gallon, so

novelties by tier and differ only in where private label sits, and I name them accordingly. Private label holds 38 percent of volume, but its similarity to Blue Bell is not captured by the seven recorded characteristics. The two groupings allow this similarity to be represented differently. The *PL-separate* grouping gives private label a group of its own. The *PL-pooled* grouping places private label in the group of mainstream tubs. Table 6 lists the groups of each grouping and every specification by name; the nested logits use these groups as nests, and a variant of the PL-separate grouping that keeps the novelties in one nest is reported in Appendix B. Blue Bell’s volume is concentrated in mainstream tubs (89 percent of its May–August 2014 gallons in its markets), so the groupings differ mainly in whether private label shares Blue Bell’s group. RCNL and the group RC use the PL-pooled groups, and each product keeps its group in the counterfactuals. Because RCNL uses the same nests, its result is not independent evidence for the grouping, and the PL-separate nested logit provides the comparison in which private label forms its own nest.

The grouping’s two dimensions follow how the industry describes the category, by format and price tier (U.S. Department of Agriculture, Agricultural Marketing Service, 2003; Federal Trade Commission, 2003; Kennedy, 2013; Draganska et al., 2009; Sullivan, 2020).⁵ Placing private label among the mainstream tubs is my classification, based on its price, not an industry label. It is consistent with evidence that retailers position store brands against the leading national brand, with which they compete most closely in categories where store-brand quality is high (Sayman et al., 2002). No grouping singles out a brand, and every grouping is defined from pre-withdrawal prices and formats.

The group RC uses the same product groups but represents their similarity through persistent group tastes rather than a shared within-period shock. In the group RC,

$$\eta_{ig} = \sigma_g z_{ig}, \quad z_i \sim N(0, I), \quad (9)$$

so each group has its own dispersion of tastes, where RCNL imposes one ρ on its nests. Although the group components are independent, a consumer’s utilities for products in different groups stay correlated through the characteristic and inside-good tastes they share, because products carry several characteristics at once and a characteristic appears in several groups.

any cutoff in that range gives the same groups. Novelty prices are dense around \$25, so that cutoff is a round number rather than a break in the price distribution.

⁵Government purchase specifications distinguish economy, standard, premium, and superpremium ice cream (U.S. Department of Agriculture, Agricultural Marketing Service, 2003), antitrust practice separates superpremium from premium and economy ice cream (Federal Trade Commission, 2003), earlier studies of the industry use similar tiers (Draganska et al., 2009; Sullivan, 2020), and syndicated retail data report ice cream and frozen novelties as separate categories (Kennedy, 2013). My two tiers aggregate these finer segments by price.

4.2 Substitution after a whole-brand removal

The withdrawal poses two questions, and each specification answers them differently. Let $\mathcal{B}$ be the set of Blue Bell products, $s_{\mathcal{B}}$ their combined share, and s_0 the outside share in a market. Removal diversion to a destination k is the increase in k 's quantity per unit of Blue Bell's lost quantity, holding prices and the other products' mean utilities fixed:

$$D_{\mathcal{B},k}^R = \frac{q_k(\mathcal{J} \setminus \mathcal{B}) - q_k(\mathcal{J})}{\sum_{j \in \mathcal{B}} q_j(\mathcal{J})}. \quad (10)$$

Retention is its sum over the inside destinations, measured per unit of Blue Bell's lost quantity. Allocation is each destination's share of the demand retained by inside products, $D_{\mathcal{B},k}^R / \sum_{k'} D_{\mathcal{B},k'}^R$. Section 6.1 normalizes the removal effects by Blue Bell's 2014 gallons.

In logit, the removed products' demand is reallocated in proportion to the remaining shares:

$$D_{\mathcal{B},k}^R = \frac{s_k}{1 - s_{\mathcal{B}}}, \quad R_{\mathcal{B}} = 1 - \frac{s_0}{1 - s_{\mathcal{B}}}. \quad (11)$$

Retention is fixed by the outside share and allocation by the destinations' market shares, and neither depends on who bought Blue Bell. Logit leaves no room for the identity of Blue Bell's buyers to shape where their demand goes.

In nested logit, the products that remain in a nest that loses Blue Bell products gain proportionally more than products in other nests. Let $s_{\mathcal{B}|g}$ be the removed products' share of nest g and s_g the nest's share, with $s_{\mathcal{B}|g} = 0$ for nests without Blue Bell products. A remaining product k in nest $g(k)$ gains by the factor

$$\frac{s'_k}{s_k} = \frac{(1 - s_{\mathcal{B}|g(k)})^{-\rho}}{\Delta}, \quad \Delta = 1 - \sum_g s_g \left[1 - (1 - s_{\mathcal{B}|g})^{1-\rho} \right], \quad (12)$$

and the outside share rises to s_0/Δ . With $\rho = 0$ every remaining product gains by the factor $1/(1 - s_{\mathcal{B}})$, the logit case. As ρ approaches one, Δ approaches one and the removed demand stays within its nests. Because the outside option is a singleton nest, a larger ρ raises retention as well as concentrating allocation within nests: nested logit ties the two margins to one parameter.

In the non-nested RC specifications, removal diversion averages individual logit diversion over consumers, weighting each consumer by its probability of buying Blue Bell. Writing $P_{i\mathcal{B}} = \sum_{j \in \mathcal{B}} P_{ij}$,

$$D_{\mathcal{B},k}^R = \frac{\int P_{i\mathcal{B}} P_{ik} / (1 - P_{i\mathcal{B}}) dF(i)}{\int P_{i\mathcal{B}} dF(i)}, \quad R_{\mathcal{B}} = 1 - \frac{\int P_{i\mathcal{B}} P_{i0} / (1 - P_{i\mathcal{B}}) dF(i)}{\int P_{i\mathcal{B}} dF(i)}. \quad (13)$$

The weights select Blue Bell’s buyers. Dispersion in the inside-good taste is one channel through which this selection raises retention: consumers with a high taste $\sigma_0\nu_{i0}$ have a small outside probability P_{i0} , and Blue Bell’s buyers are selected toward them. Tastes for characteristics direct the retained volume toward products that share Blue Bell’s characteristics, and the income terms move both margins through price sensitivity and the level of utility of inside goods. Two RC specifications with the same market shares can hence differ on both margins, depending on how their estimated tastes describe the consumers who chose Blue Bell. Group-specific random coefficients add a persistent taste for each product group, so Blue Bell’s buyers are selected on their group tastes and their characteristic tastes jointly, and a dispersed taste for Blue Bell’s group keeps its buyers within that group when it leaves.

4.3 Identification and micro moments

Prices are endogenous, so I instrument for price. The excluded instruments, in the spirit of Berry et al. (1995), are the sums of the characteristics of the firm’s other products and of rival products in the market, the number of rival brands, and the cane sugar price and its interactions with the vanilla, chocolate, and cookies indicators. The nested logits and RCNL add the within-nest instruments, the counts and characteristic sums of the products in the same nest (Berry, 1994; Verboven, 1996). RCNL omits the sugar-price instruments, with which its implied elasticities lie well outside the range of the other specifications (Appendix B). These instruments are valid if, after the fixed effects, sugar prices and the characteristics and availability of rival products are unrelated to the unobserved demand term. All parameters are estimated jointly. The price coefficient α is identified by price variation within products and markets over time, after the fixed effects, using these instruments; the micro moments discipline the taste distribution and add information that does not rely on the excluded instruments, which matters because weak instruments can make the estimated nonlinear parameters unreliable (Conlon and Gortmaker, 2020; Gandhi and Houde, 2020).

The micro moments describe annual purchases of the Homescan households of 2011 through 2014. Most are conditional annual quantities: the mean number of gallons bought in a year by the households that buy a given set of products, in gallons per household-year. A household counts as a buyer of a set when it buys at least one gallon of it in the year; in the model, each simulated consumer chooses once a month, and the annual moments are built from its twelve monthly choice probabilities (Appendix D). The moments fall into three blocks, and each carries different information about the consumers behind aggregate shares.

The base block, targeted by every RC specification, has 12 moments and follows Fan and Yang (2025). Eight are conditional annual quantities: the mean annual quantity of the

households that buy ice cream in the year, and the mean annual quantity of the products with each of the seven characteristics among the households that buy such products. Persistent taste heterogeneity concentrates repeated purchases among households with stronger preferences for a characteristic. Conditional annual quantities therefore provide information about taste dispersion beyond aggregate shares: characteristic-specific quantities inform the corresponding tastes, and total ice cream purchases among buyers inform the inside-good taste. The other four, similar to those of Petrin (2002), are the differences between the middle or high and the low income group in quantity and in the price paid: the price differences discipline γ_p , and the quantity differences the income shifts defined on the same three groups. The inside-good standard deviation needs this information: with market fixed effects, unobserved heterogeneity in the preference for the outside good is not identified from aggregate shares alone (Conlon and Gortmaker, 2025), and it governs substitution to the outside good (Conlon and Mortimer, 2021). Through equation (13), the inside-good and characteristic tastes set how strongly Blue Bell’s buyers are selected toward the category, the channel through which the RC specifications can raise retention. The eight income-group contrasts, targeted only by the demographic extension, are the differences between income groups in the volume shares of premium products, novelties, and private label and in the price paid.

The four own-group moments are the annual purchases within a group among the households that buy the group, the group counterpart of the conditional quantities above; the group RC targets them on its PL-pooled groups. They measure how persistently households return to a group over the months of a year. Persistent group and category tastes can generate correlation across purchases; a nest shock redrawn independently each month does not itself generate persistence. Moments of this kind, which relate the products a consumer chooses on different occasions with tastes fixed across occasions and choice shocks drawn anew, are the consumer-panel moments that Berry and Haile (2021) describe. They discipline the group standard deviations of the group RC directly, and a dispersed taste for Blue Bell’s group would keep its buyers within that group when it leaves. RCNL does not target them: its nesting parameter is determined by the scanner moments with the within-nest instruments together with the base micro moments, and first-choice demographic moments carry little information about nesting (Conlon and Gortmaker, 2025).

4.4 Estimation

Let n be the number of product–market–month observations, Z_{jmt} the instruments, $H(\theta)$ the model’s micro moments, and $\widehat{H}$ their observed values, and define

$$g(\theta) = \frac{1}{n} \sum_{jmt} Z_{jmt} \xi_{jmt}(\theta), \quad d(\theta) = H(\theta) - \widehat{H}. \quad (14)$$

The estimator minimizes

$$Q(\theta) = n g(\theta)' S_A^{-1} g(\theta) + d(\theta)' V_M^{-1} d(\theta), \quad (15)$$

which is two-step GMM. S_A is the covariance of $Z\xi$, clustered two ways by product and by DMA, evaluated at a first-step estimate and held fixed during optimization; V_M is the covariance of the micro-moment targets across whole-household bootstrap resamples (Appendix B). The scanner and household blocks come from separate samples and are treated as independent.

5 Demand Estimates

Table 7 reports the estimated price sensitivity, taste dispersion, and nesting parameters. I use these estimates to interpret the models’ retention and allocation predictions below. Appendix Tables B.2 and B.3 report the nested logit with novelties in one nest, the RC without retailer-by-private-label effects, the demographic extension, and the estimates on the longer 2011–2016 window.

The two nested logits have nesting parameters of 0.77 (PL separate) and 0.82 (PL pooled) and nearly the same price coefficient. Each is estimated with its own nests and within-nest instruments, but their similar price sensitivity and strength of nesting make the composition of the nests a useful explanation of the differences between them.

The RC specifications differ most in the dispersion of the taste for the category as a whole. The inside-good standard deviation is 0.92 in the RC model, with or without the retailer-by-private-label effects, and 0.98 in the demographic extension. In RCNL it is 1.19 (standard error 0.07), and in the group RC it is 0.02 (standard error 1.47). Its scale is that of the type-I extreme-value shock, whose standard deviation is $\pi/\sqrt{6} \approx 1.28$, so in the RC model the persistent category taste varies across consumers about seven tenths as much as the idiosyncratic shock of a single choice. Dispersion in the category taste can raise retention by selecting Blue Bell buyers who also value the remaining inside products; Section 6.3 examines how strongly this selection operates at the estimated parameters.

Table 7: Demand estimates of the six specifications

	NL, Logit PL separate		NL, PL pooled		RC	RCNL	Group RC
<i>Price</i>							
Price coefficient, α	−0.211 (0.013)	−0.098 (0.007)	−0.093 (0.006)	−0.479 (0.034)	−0.233 (0.022)	−0.443 (0.035)	
Price $\times$ log income, γ_p	.	.	.	0.0126 (0.0026)	0.0085 (0.0020)	0.0132 (0.0025)	
<i>Substitution structure</i>							
Nesting parameter, ρ	.	0.769 (0.013)	0.823 (0.013)	.	0.718 (0.028)	.	
Inside-good SD, σ_0	.	.	.	0.925 (0.084)	1.194 (0.072)	0.018 (1.468)	
Mainstream tubs and PL SD	.	.	.	.	.	0.033 (2.265)	
Premium tubs SD	.	.	.	.	.	0.043 (0.933)	
Mainstream novelties SD	.	.	.	.	.	0.563 (0.169)	
Premium novelties SD	.	.	.	.	.	0.095 (0.748)	
<i>Characteristic tastes</i>							
Vanilla SD	.	.	.	0.657 (0.038)	0.204 (0.027)	0.483 (0.092)	
Chocolate SD	.	.	.	0.775 (0.030)	0.269 (0.026)	0.719 (0.070)	
Bulk SD	.	.	.	0.911 (0.029)	0.374 (0.035)	0.903 (0.214)	
Low-fat SD	.	.	.	0.852 (0.140)	0.251 (0.089)	0.691 (0.201)	
Low-calorie SD	.	.	.	0.830 (0.050)	0.302 (0.036)	0.663 (0.162)	
Assorted SD	.	.	.	0.017 (2.153)	0.003 (1.705)	0.054 (7.313)	
Cookies / mix-ins SD	.	.	.	0.084 (7.296)	0.014 (3.295)	0.053 (9.513)	
<i>Income shifts</i>							
Middle income, λ_{mid}	.	.	.	0.031 (0.063)	0.232 (0.044)	−0.123 (0.266)	
High income, λ_{high}	.	.	.	−0.560 (0.069)	−0.490 (0.055)	−0.369 (0.456)	
Own-price elasticity, gallon-weighted median	−1.87	−3.41	−4.51	−2.99	−4.20	−2.64	
<i>Fixed effects</i>							
Product, DMA–retailer, year, month	Yes	Yes	Yes	Yes	Yes	Yes	
Retailer $\times$ private label	No	Yes	Yes	Yes	No	No	
Instruments	Base	Base and within-nest	Base and within-nest	Base	Base and within-nest, no sugar prices	Base	
Micro moments	None	None	None	Base	Base	Base and own-group	
Product–market–months				3,705,958			
Households	.	.	.	84,609	84,609	84,609	

Standard errors in parentheses: clustered by DMA (117 clusters) for logit and the nested logits, GMM standard errors for the RC specifications. Units: α per dollar per gallon, the intercept of the RC price coefficient; γ_p per unit of log income; income shifts relative to the low income group. A dot marks an excluded parameter. Sanderson–Windmeijer first-stage F statistics for price and within-nest share: 363 and 362 (PL separate), 357 and 275 (PL pooled). Elasticities cover rival product-market-months of May–August 2014.

The RC model places substantial taste dispersion on five of the seven characteristics, vanilla, chocolate, bulk, low fat, and low calorie. For assorted products and cookies or mix-ins the standard deviations are estimated near zero but imprecisely.

RCNL combines a nesting parameter of 0.72 (standard error 0.03) with category and characteristic tastes. Its characteristic tastes also shape allocation within nests. Its raw characteristic standard deviations are not comparable with the RC model's. Choice within a nest depends on $v_{ijmt}/(1-\rho)$ (equation (8)), so within a nest a taste with standard deviation σ acts like one with standard deviation $\sigma/(1-\rho)$ in the RC model, and at the estimated ρ the factor $1/(1-\rho)$ is 3.55. On that scale RCNL's characteristic tastes within a nest are about as dispersed as the RC model's or slightly more; bulk, for example, has a scaled standard deviation of 1.33 against 0.91.⁶ Across nests and toward the outside option, where the scale is that of the full extreme-value shock, its standard deviations for the five characteristics that carry dispersion in the RC model are 0.20 to 0.37, and those for assorted products and cookies or mix-ins are estimated near zero.

The group RC's estimated dispersion is concentrated on one group, mainstream novelties, with a standard deviation of 0.56 (standard error 0.17). For the group that holds Blue Bell's mainstream tubs and private label it is estimated near zero but imprecisely, at 0.03, as are the standard deviations of the two premium groups. For the five characteristics that carry dispersion, its standard deviations lie between 0.48 and 0.90, below the RC model's except for bulk, which is about the same.

Income moves both price sensitivity and the taste for the category. The interaction γ_p is positive in every specification, between 0.008 and 0.013 per unit of log income, so higher-income consumers respond less to price. It is 0.013 (standard error 0.003) in both the main RC and the group RC, and 0.008 (0.002) in RCNL. In the RC model the middle-income shift is close to zero and the high-income shift is -0.56 (standard error 0.07); in RCNL they are 0.23 (0.04) and -0.49 (0.05). The high-income shift is negative: at a given price it lowers the utility of inside goods for high-income households, while their smaller price sensitivity raises it. In the group RC both income shifts are imprecisely estimated (standard errors 0.27 and 0.46).

Adding the income-group contrasts changes the key parameters little; estimating on the window through 2016, which includes the withdrawal and the return, changes some of them, such as the RC model's inside-good standard deviation, from 0.92 to 1.07 (Appendix Table B.3).

In the RC specifications, the price coefficient rises with income but remains negative for

⁶The scaled standard deviations are 0.72 for vanilla, 0.96 for chocolate, 0.89 for low fat, and 1.07 for low calorie, against 0.66, 0.78, 0.85, and 0.83 in the RC model.

every simulated consumer. The gallon-weighted median own-price elasticity over the rival products of May–August 2014 is -3.0 in the main RC, against -1.9 in logit. On the same four nests the nested logit, with ρ of 0.82, has an elasticity of -4.5 , and RCNL, with ρ of 0.72, one of -4.2 . A product that raises its price sends most of its lost sales to other inside products, mainly those in its own nest, and only 0.13 of them to the outside option in RCNL and 0.10 in the PL-pooled nested logit, against 0.41 in the RC model (Appendix Table E.1). This outside share of a marginal price response should not be read as outside diversion from removing the whole brand; RCNL’s withdrawal retention is 0.73 (Section 6.2). The outside share is the marginal response to one product’s price with its nest intact, and the lost sales go mostly to the other products of the nest, Blue Bell’s own among them. The withdrawal removes every Blue Bell product at once, and by equation (12) the share of removed demand that leaves the nest rises with the removed products’ share of their nest. The price response and the withdrawal also weight consumers differently: the price response depends on consumers near the margin of switching, and the removal response on all of Blue Bell’s buyers (Conlon and Mortimer, 2021).

Appendix D reports the available micro-moment fits.

6 Predictive Performance

6.1 Evaluation design

The comparison evaluates substitution at the observed 2015 rival prices and sales. For each DMA–retailer market and month of May–August 2015, I invert each model’s demand system to recover rival mean utilities that reproduce observed sales without Blue Bell (Berry et al., 1995). I then return Blue Bell at its forecast mean utility and its last first-quarter 2015 price. The gallons its presence would have taken from a group d of rivals are the model’s removal effect,

$$M_{dmt} = \sum_{j \in d} \left(q_{jmt}^{obs} - q_{jmt}(\tilde{\delta}_{mt}, \hat{\delta}_{B,mt}; \mathcal{J}_{mt}) \right), \quad (16)$$

where $\tilde{\delta}_{mt}$ are the rivals’ recovered mean utilities and $\hat{\delta}_{B,mt}$ Blue Bell’s forecast utility, which carries forward its product, market, and time effects and its local demand in the months before the complete withdrawal under the same rule for every model (Appendix C gives the rule). Every rival keeps its observed 2015 price, so no model is asked to predict how prices responded to the withdrawal, and whatever else changed in a market after the recall sits in the rivals’ recovered utilities and stays there when Blue Bell returns. The evaluation therefore asks a conditional question: given the rival products as consumers found them

in the summer of 2015, at the prices and with the appeal they had, how would sales have divided had Blue Bell also been on the shelf at its last pre-recall price? The answer is removal diversion evaluated at the fixed 2015 market state, in the sense of Conlon and Mortimer (2021). Raval et al. (2022) also evaluate removal predictions, for hospital closures with the remaining hospitals as they were. It is not an equilibrium counterfactual in which rivals reprice: how rival prices responded to the withdrawal is a supply-side question, tested separately in Section 7, and the sales generated by that price response are part of the second term of equation (18).

A removal effect is the product of how much Blue Bell volume the model has in the market-month and where that volume goes. I rescale each market-month’s effect to its 2014 same-month Blue Bell gallons V_{mt} ,

$$M_{dmt}^a = M_{dmt} \frac{V_{mt}}{\widehat{B}_{mt}}, \quad (17)$$

where $\widehat{B}_{mt}$ is the model’s Blue Bell gallons with Blue Bell present, so that market-months have the same volume weights across models and differences in forecast Blue Bell volume do not mechanically scale the removal effects. The rescaling leaves Blue Bell’s utility, and with it where its demand goes in the nested and random-coefficient models, as it was.

Products never sold anywhere by March 2015, new listings among them, and products whose 2015 price fails the checks of Appendix C are held at their observed gallons, 2.2 percent of rival gallons in Blue Bell’s markets (Appendix Table C.1). Only the scored products and the outside option respond to Blue Bell’s return; the other products remain at their observed quantities.

The observed change in rival sales may also reflect changes the recall caused in prices, promotions, or tastes. Let $q_R(A, x)$ be the sales of a group of rivals when Blue Bell is available ($A = 1$) or not ($A = 0$) and the rest of the market is in state x , which collects prices, displays, promotions, assortment, and tastes. With x_1 the state in 2015 and x_0 the state 2015 would have had without the recall,

$$q_R(0, x_1) - q_R(1, x_0) = [q_R(0, x_1) - q_R(1, x_1)] + [q_R(1, x_1) - q_R(1, x_0)]. \quad (18)$$

The observed effect of Section 3.1 estimates the left side, and the design predicts the first term on the right, the effect of Blue Bell’s absence in the 2015 market. The second term, the change in rival sales that the recall caused through other channels, moves the benchmark for every model alike; Section 3 examines pre-recall comparability and gives the household evidence, Section 7 documents the rival price changes, and Appendix F evaluates how a

price-induced sales gain would alter the comparison.

Model and observed effects are computed on the same product–markets, those that sold in the same month of 2014, grouped into six destinations for every model, Blue Bunny, private label, Breyers, Dreyer’s, premium pints (Ben & Jerry’s and Häagen-Dazs), and other brands, and summed over the 1,246 market–months $\mathcal{R}$ of Section 2:

$$m_d = \frac{\sum_{mt \in \mathcal{R}} M_{dmt}^a}{\sum_{mt \in \mathcal{R}} V_{mt}}, \quad o_d = \frac{\sum_{mt \in \mathcal{R}} O_{dmt}}{\sum_{mt \in \mathcal{R}} V_{mt}}, \quad (19)$$

where O_{dmt} is the observed effect of equation (1). The retention gap T asks whether the model keeps the right amount of Blue Bell’s demand with the rivals, the allocation error C whether it sends the retained demand to the right rivals, and the misplaced share L combines the two:

$$T = \sum_d m_d - \sum_d o_d, \quad (20)$$

$$C = \frac{1}{2} \sum_d \left| \frac{m_d}{\sum_k m_k} - \frac{o_d}{\sum_k o_k} \right|, \quad (21)$$

$$L = \frac{1}{2} \left(\sum_d |m_d - o_d| + |T| \right). \quad (22)$$

The retention gap is the model’s rival gain minus the observed one, per 2014 Blue Bell gallon, negative when the model’s rivals gain less than the observed ones. The allocation error is the share of the rival gain assigned to the wrong rivals. The misplaced share combines errors across the six rival destinations with the error in the remaining share not accounted for by those destinations. In the model this remainder is the outside option; in the observed benchmark it also includes gains to unscored products and errors in the estimated trend. When both vectors are nonnegative and sum to one, the score is the smallest share of Blue Bell’s volume that must be reassigned to reconcile the model’s allocation with the observed one (Appendix C works through an example). Smaller is better for all three, for the retention gap in absolute value. The scores compare gains pooled across market–months, so overprediction in one market–month can offset underprediction in another (Appendix C). Intervals come from a block bootstrap over DMAs with the parameter estimates held fixed, so they reflect the sampling of markets and not estimation error, with 199 draws for intervals, 399 for paired comparisons, and 499 for the smaller market size of Section 8. The draws are common to all models, so comparisons between models are paired, and the DMAs share one recall, so the intervals describe the dispersion of the evaluation panel around the recall that occurred, not variation across recalls.

6.2 Retention and allocation

The two margins rank the models differently. The nested specifications come closer to the observed retention, but nesting alone does not improve allocation: logit, the RC model, and the group RC allocate their smaller retained volumes about as accurately as the PL-pooled nested logit (Table 8 and Figure 5). On retention, the nested specifications keep most of Blue Bell’s demand with the rivals: the PL-pooled nested logit keeps 0.74 gallons of rival sales per 2014 Blue Bell gallon against 0.77 observed, RCNL 0.73, and the PL-separate nested logit 0.62. Logit, the RC model, and the group RC keep 0.42 to 0.46, about 0.3 gallons less than the observed retention.

On allocation, the PL-pooled nested logit gives private label 0.43 gallons per Blue Bell gallon against 0.40 observed, and its allocation error, 0.12, is in the range of logit’s, the RC model’s, and the group RC’s, 0.11 to 0.13. RCNL’s allocation error, 0.09, is the lowest of the six, although it is lower than those of logit, the RC model, and the group RC in only 75 to 87 percent of paired draws. The PL-separate nested logit allocates the retained demand less accurately, with an allocation error of 0.33: it closely matches Dreyer’s gain (0.16 against 0.16 observed) but substantially underpredicts private label’s gain and overpredicts Breyers’; the PL-pooled nested logit improves the joint allocation while still underpredicting Dreyer’s.

On the combined score, the two specifications with PL-pooled nests have the lowest misplaced shares, about a tenth of Blue Bell’s volume, 0.101 for the nested logit and 0.104 for RCNL, against 0.29 to 0.35 for the other four, and they are the two models close on both margins. The nested specifications’ advantage on retention is clear: each comes closer to the observed retention than logit, the RC model, and the group RC in 96 to 99 percent of paired draws. On allocation alone, the paired comparisons do not establish a clear ranking of the PL-pooled nested logit, logit, the RC model, and the group RC (Appendix Table C.4). The uncertainty in the combined ranking comes mainly from the observed retention, whose interval runs from 0.54 to 1.01. The paired difference in the misplaced share between the PL-pooled and the PL-separate nested logit is -0.19 , with a 95 percent interval from -0.27 to 0.13, and in draws where the observed retention is low the PL-separate nested logit comes closer on the level. Each of the two PL-pooled specifications has a lower misplaced share than each of the other four in 93 to 96 percent of paired draws, and the nested logit is lower than RCNL in 54 percent; these frequencies describe pairwise rankings under resampling of this event’s DMAs.

Figure 6 adds the two destinations that Table 8 leaves out, premium pints and other brands, and shows why getting retention right does not settle where the demand goes. The PL-separate nested logit sends more than the other models to the branded tubs that share Blue Bell’s nest in its grouping, which matches Dreyer’s gain but overpredicts those of Breyers

Table 8: Retention and allocation during the withdrawal

	Observed	NL, Logit PL separate		NL, PL pooled	RC	RCNL	Group RC
<i>A. Gallons gained per 2014 Blue Bell gallon</i>							
All rivals (retention)	0.774 [0.544, 1.006]	0.424 [0.399, 0.446]	0.619 [0.581, 0.648]	0.742 [0.686, 0.766]	0.463 [0.439, 0.483]	0.735 [0.690, 0.754]	0.443 [0.415, 0.467]
Private label	0.398 [0.176, 0.569]	0.216 [0.198, 0.229]	0.117 [0.096, 0.129]	0.427 [0.388, 0.449]	0.220 [0.201, 0.235]	0.361 [0.329, 0.381]	0.208 [0.190, 0.221]
Blue Bunny	0.014 [−0.009, 0.033]	0.014 [0.009, 0.022]	0.040 [0.027, 0.063]	0.027 [0.018, 0.039]	0.017 [0.012, 0.026]	0.030 [0.021, 0.045]	0.016 [0.011, 0.026]
Breyers	0.033 [−0.026, 0.059]	0.032 [0.030, 0.035]	0.115 [0.106, 0.128]	0.062 [0.057, 0.067]	0.039 [0.036, 0.044]	0.081 [0.075, 0.087]	0.039 [0.036, 0.043]
Dreyer's	0.160 [0.119, 0.246]	0.045 [0.038, 0.054]	0.159 [0.133, 0.187]	0.093 [0.078, 0.108]	0.053 [0.046, 0.064]	0.111 [0.093, 0.127]	0.053 [0.045, 0.064]
<i>B. Errors</i>							
Retention gap (model minus observed)	·	−0.350	−0.155	−0.032	−0.311	−0.040	−0.331
Allocation error (share of the rival gain)	·	0.106	0.329	0.118	0.131	0.091	0.134
Misplaced share of Blue Bell's volume	·	0.350	0.291	0.101	0.322	0.104	0.340

Evaluation design of Section 6.1 on the 1,246 market-months with 2014 Blue Bell sales. Gallons gained per 2014 Blue Bell gallon, with 95 percent DMA-bootstrap intervals in brackets; Figure 6 adds premium pints and other brands. The three scores are equations (20) to (22).

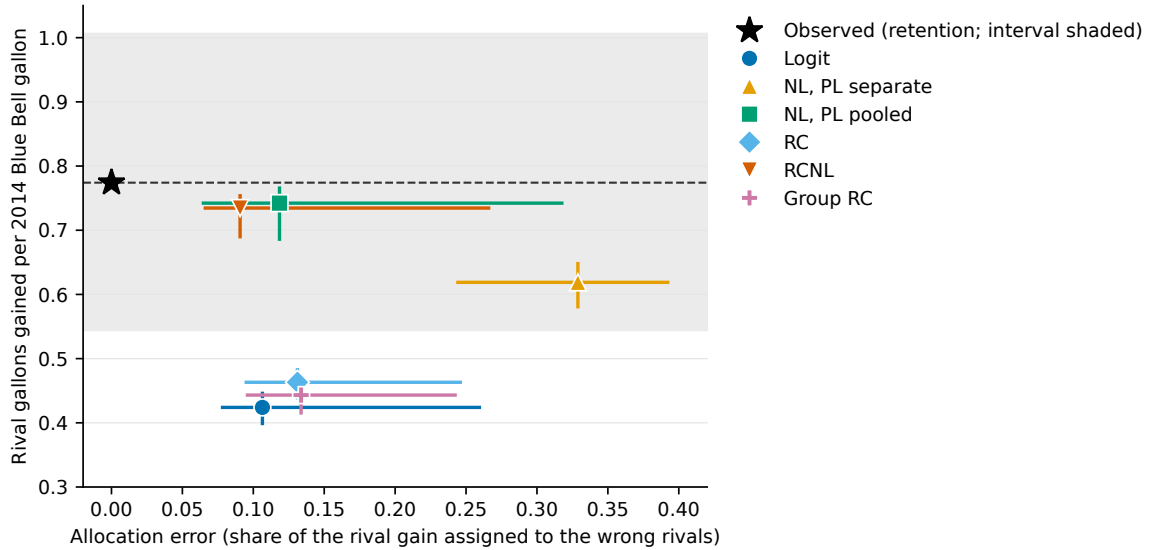


Figure 5: Retention and allocation of the six specifications

Vertical axis: rival gallons gained per 2014 Blue Bell gallon; horizontal axis: the allocation error of equation (21); bars and the shaded band are 95 percent DMA-bootstrap intervals, and the observed point has an allocation error of zero by construction.

and Blue Bunny, and too little to private label. RCNL and the PL-pooled nested logit retain about the same amount and divide it slightly differently: RCNL sends less to private label, 0.36 gallons per Blue Bell gallon against 0.43 (0.40 observed), and more to Dreyer’s, Breyers, and other brands, so it comes closer on Dreyer’s and other brands and less close on private label and Breyers. Logit, the RC model, and the group RC send about half of what they retain to private label, close to the observed share, and about a fifth to other brands, so they are competitive on conditional allocation. Their error is in the level: they give private label, Dreyer’s, and other brands less than the observed gains, Dreyer’s furthest below in proportion, and come close on the three small destinations. Premium pints gained little in the data and gain little in every model.

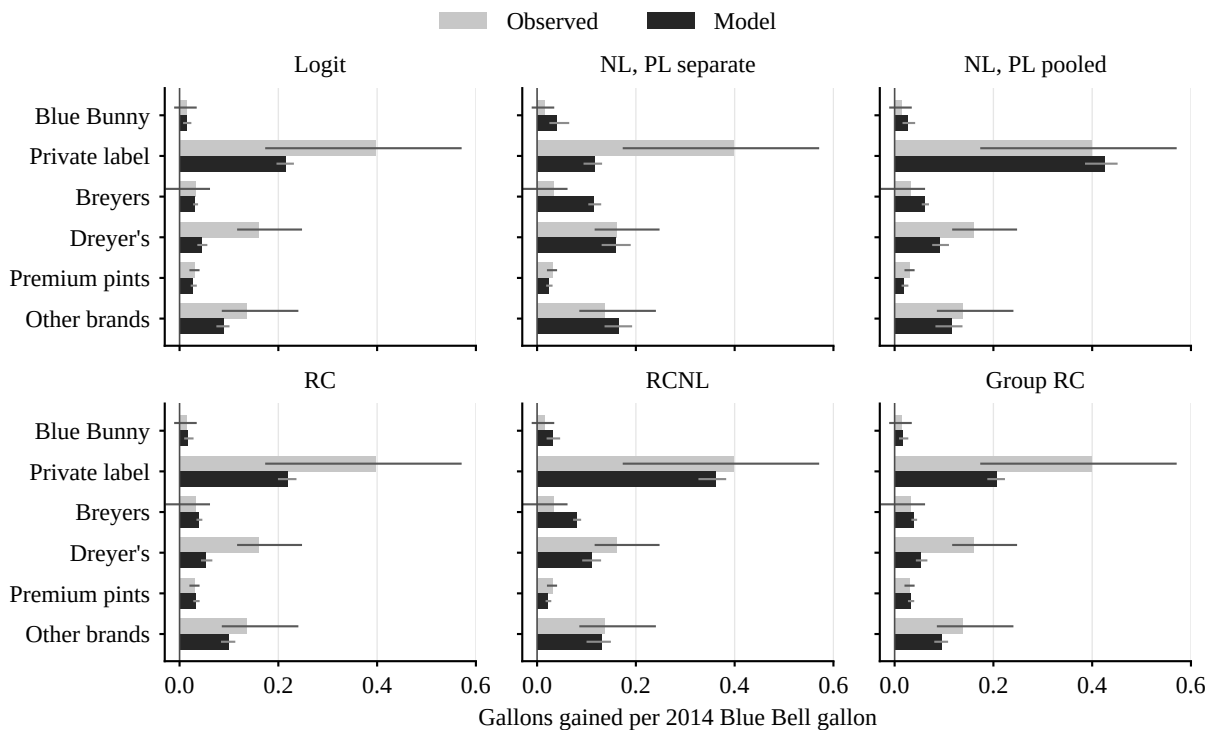


Figure 6: Gallons gained by destination

Rival gallons gained per 2014 Blue Bell gallon by destination during the withdrawal, evaluation design of Section 6.1 on the 1,246 market-months. Gray: observed; black: model; lines: 95 percent DMA-bootstrap intervals. Premium pints are Ben & Jerry’s and Häagen-Dazs.

6.3 Interpreting the prediction errors

Logit’s retention is set by the outside share. By equation (11), the share of Blue Bell’s demand that stays inside is $1 - s_0/(1 - s_B)$, one minus the outside share of the market

without Blue Bell, which in the evaluation design is the observed 2015 market. The logit prediction is thus the IIA restriction evaluated at the chosen market size, and it depends on nothing the model estimates; Section 8 examines sensitivity to a smaller market size. Logit’s allocation is proportional to the rivals’ sales in each market–month, which gives private label about the half of the gain it took and Dreyer’s about half of its share of the gain (0.11 against 0.21).

The RC model can keep more only through the selection of Blue Bell’s buyers. By equation (13), its retention exceeds logit’s to the extent that the consumers who bought Blue Bell have smaller outside probabilities than those who did not. Two estimated tastes produce that selection: the inside-good taste, which raises the utility of every inside product for some consumers, and the characteristic tastes, which raise the utility of products with Blue Bell’s characteristics. At the estimated parameters the RC model retains only 0.04 more than logit: the category taste makes a consumer’s utilities for inside products positively dependent, but the resulting selection of Blue Bell’s buyers leaves it far short of the observed retention. As a description of that dependence, the inside-good taste ($\sigma_0 = 0.92$) gives two inside products a utility correlation of about 0.34, whereas two products in the same nest of a nested logit have a correlation above 0.94.⁷ The correlations describe the dependence each model allows; they do not decompose the difference in retention.

The nested logits keep demand inside through the nest. By equation (12), with ρ near 0.8 most of Blue Bell’s demand stays among the products of its own nest, so retention is high in both groupings, and allocation follows which products share Blue Bell’s nest. In the PL-separate grouping, Blue Bell’s mainstream-tub nest contains the branded mainstream tubs and not private label, so the model sends the retained demand to Breyers, Dreyer’s, and Blue Bunny and gives private label about a fifth. In the PL-pooled grouping, private label is in that nest and receives about the share it took. The two groupings differ only in where private label sits. The PL-separate grouping and its variant with novelties in one nest give nearly the same misplaced shares, 0.291 and 0.287 (Tables 8 and 9), so the novelty split makes little difference to the PL-separate specification’s combined error. The PL-pooled grouping allows strong substitution between Blue Bell and private label without requiring them to share recorded characteristics. This matters because Nielsen masks private-label characteristics

⁷With an inside-good standard deviation of 0.92 and type-I extreme-value shocks of variance $\pi^2/6$, two inside products with no characteristic that carries a random taste have a utility correlation of $0.92^2/(0.92^2 + \pi^2/6) = 0.34$ (Train, 2009, Section 6.3). Two vanilla products in bulk containers, which also share the vanilla and bulk tastes, have about 0.56, and the income terms add little; a product with further characteristics has a larger utility variance and a lower correlation with the other. In a nested logit, by equation (5), the nest component has variance $[1 - (1 - \rho)^2]\pi^2/6$, which is the covariance of the shocks of two products in the same nest, so their correlation is $1 - (1 - \rho)^2$ (Galichon, 2022, Proposition 1), 0.95 to 0.97 at the estimated nesting parameters. A correlation of 0.95 would need σ_0 of about 5.6.

(Section 2), so private-label products carry none of the seven in the estimation data, and random tastes for those characteristics cannot directly represent their similarity to Blue Bell, although the category taste and the income terms can still link their demands. Removing the whole brand also removes Blue Bell products that would otherwise substitute for one another. In nested logit, equation (12) shows that a larger removed share of the nest sends more demand outside it; the same expression applies within each consumer type in RCNL (Section 5).

The grouping also moves retention. For a given Blue Bell share and nesting parameter, a larger nest around Blue Bell lowers the outside option’s gain from the removal, because more of the demand that leaves Blue Bell finds a close substitute within its nest (Appendix B derives this from equation (12)). Adding private label, about half of rival sales, to Blue Bell’s nest is consistent with the PL-pooled nested logit keeping 0.74 against 0.62 for the PL-separate one, whose nesting parameter is also lower (0.77 against 0.82).

With the same PL-pooled nests, RCNL and the nested logit perform similarly on both margins. RCNL keeps 0.73 against the nested logit’s 0.74, and the nested logit is closer to the observed retention in 63 percent of paired draws; RCNL’s allocation error is 0.09 against 0.12, lower in 67 percent; and their misplaced shares are nearly equal, 0.104 and 0.101. Characteristic tastes allow RCNL to favor products that share Blue Bell’s recorded attributes, which private label lacks in the data. This is one possible explanation for its smaller predicted gain for private label, but the comparison does not isolate that channel.

The group RC asks whether the same groups without nesting reproduce the nested models’ high retention. At the estimated parameters they do not: the standard deviation of the taste for Blue Bell’s group is 0.03, imprecisely estimated, and retention is 0.44, close to the RC model’s. Group random coefficients represent persistent differences in tastes for a group, which act on purchases over the whole year as well as on diversion after a product disappears, whereas the nest adds a shock shared by the products of a group within a single choice. Using the same groups therefore does not make the two specifications equivalent.

6.4 Specification comparisons

Table 9 compares specifications in two ways. Panel A compares specifications with and without additional mean-utility controls or micro moments, keeping the other specification choices unchanged. Holding the nonlinear parameters fixed, removing the retailer-by-private-label effects from the RC model changes the removal predictions by at most 0.002. Because rival mean utilities are recovered from observed sales, a mean-utility control can affect this comparison only through Blue Bell’s forecast utility when the nonlinear parameters are held

fixed. When the RC model is re-estimated without these effects (Appendix Table B.3), its nonlinear parameters change little, apart from two imprecise standard deviations. The other two matched comparisons, the eight income-group contrasts of the demographic extension, which is re-estimated, and retailer-by-private-label effects added to the year-effects design of the nested logit with novelties in one nest, also move the allocation error and the retention gap by at most 0.002 and the misplaced share by at most 0.001.

Table 9: Specification comparisons and predictive performance

Specification vs comparator	Also differs in	Misplaced share, spec.	Misplaced share, comp.	Change in allocation error	Change in absolute retention gap	Spec. better, share
<i>A. Matched comparisons: controls or micro moments added</i>						
RC vs RC without retailer × PL effects	Nothing else	0.322	0.323	0.001	−0.001	0.99
RC, demographic, vs RC	Nothing else	0.321	0.322	−0.002	0.000	0.72
NL, PL separate, novelties in one nest, vs the same with year effects only	Nothing else	0.287	0.288	−0.002	0.001	.
<i>B. Comparisons of estimated specifications</i>						
RCNL vs NL, PL pooled	Random tastes, micro moments, no sugar-price instruments, retailer × PL effects	0.104	0.101	−0.028	0.008	0.46
RCNL vs RC without retailer × PL effects	Nests, within-nest instruments, no sugar-price instruments	0.104	0.323	−0.040	−0.273	0.95
Group RC vs RC without retailer × PL effects	Group tastes, own-group micro moments	0.340	0.323	0.003	0.019	0.01
NL, PL pooled vs NL, PL separate, novelties in one nest	Grouping, within-nest instruments	0.101	0.287	−0.205	−0.126	0.93

Evaluation design of Section 6.1. The misplaced share and the allocation error are shares; the change in the absolute retention gap is in gallons per 2014 Blue Bell gallon. Changes are the specification minus its comparator, so a negative change is an improvement. The last column is the share of paired DMA-bootstrap draws in which the specification has the lower misplaced share; a dot marks a pair without paired draws. Panel A adds mean-utility controls or micro moments to a specification and holds the rest fixed. Panel B compares specifications as estimated, which differ in the features listed and are not experiments that change one feature: RCNL and the group RC have no retailer-by-private-label effects, and no PL-pooled nested logit exists without them. The specification without retailer-by-private-label effects is evaluated at the main RC's nonlinear parameters. The last comparison uses the PL-separate grouping that keeps the novelties in one nest (Appendix Table B.1); against NL, PL separate the change in the misplaced share is −0.189.

Panel B compares specifications as estimated. They differ in more than one feature, so its rows are comparisons of estimated specifications rather than experiments that change one feature at a time. The differences are larger, and only the comparisons that add the PL-pooled nests improve both margins: moving private label into Blue Bell’s nest, which changes the grouping and its within-nest instruments, and adding the nests to the RC model. RCNL adds random tastes and the base micro moments to the PL-pooled nested logit, omits the sugar-price instruments, and has no retailer-by-private-label effects; its allocation error is 0.03 lower and its absolute retention gap 0.01 larger, and its misplaced share is lower in 46 percent of draws. Against the RC model without retailer-by-private-label effects, RCNL, which adds nests and within-nest instruments and omits the sugar-price instruments, has a retention gap 0.27 smaller in absolute value and an allocation error 0.04 smaller, so its misplaced share is 0.22 lower, and lower in 95 percent of draws. The group RC has a higher misplaced share than the RC model without retailer-by-private-label effects in 99 percent of draws (Section 6.3).

The same micro-moment targets can accompany different removal predictions. RCNL and the RC model without retailer-by-private-label effects both target the base block, yet retain 0.73 and 0.46, respectively. Their nesting structures and instruments differ, so the comparison does not attribute the retention difference to the micro moments. Raval et al. (2022) find a related gap in hospital demand: models that predict individual choices better do not predict aggregate diversion better, whereas a random coefficient on travel time improves the diversion predictions, also when its dispersion is chosen from the other disasters’ post-closure data. In NielsenIQ soft-drink data, Conlon and Gortmaker (2025) find that demographic micro moments leave their model’s predicted diversion to the outside good after a non-diet brand’s removal at 0.93, against 0.92 without them and 0.16 among surveyed buyers, and the prediction moves only when the survey’s diversion answers are themselves matched.

6.5 The direct forecast

The comparison would look different if the models were scored on a direct forecast of 2015 sales, which mixes their substitution patterns with their forecasts of rivals’ 2015 sales levels and of Blue Bell’s own volume. Table 10 reports each model’s misplaced share at three steps. The first is the direct forecast: each model forecasts every product’s 2015 sales with and without Blue Bell, at observed 2015 rival prices and Blue Bell’s last first-quarter price, and the removal effect is the difference. The second rescales each market-month’s effect to Blue Bell’s 2014 gallons, as in equation (17), which removes the scaling by the model’s own

forecast of Blue Bell’s volume. The third also replaces the forecast of the rivals’ 2015 sales with their observed values, by the inversion of Section 6.1, so the models are compared at the same observed prices and rival quantities, with the rival mean utilities recovered separately under each demand system; this is the evaluation design.

Table 10: Misplaced share at the three steps from the direct forecast to the evaluation design

	Logit	NL, PL separate	NL, PL pooled	RC	RCNL	Group RC
<i>A. Misplaced share of Blue Bell's volume</i>						
Direct forecast	0.467	0.380	0.253	0.544	0.338	0.497
Rescaled to 2014 Blue Bell gallons	0.404	0.347	0.145	0.413	0.125	0.400
Rival sales from the data (evaluation design)	0.350	0.291	0.101	0.322	0.104	0.340
<i>B. Direct forecast</i>						
Blue Bell volume, share of 2014 gallons	0.836	0.775	0.669	0.616	0.637	0.730
Retention	0.308	0.568	0.549	0.229	0.476	0.280

Panel A: the misplaced share of equation (22). First step: direct forecast with observed 2015 rival prices and Blue Bell at its last first-quarter 2015 price. Second step: each market-month's removal effect rescaled to Blue Bell's 2014 same-month gallons. Third step: rival 2015 sales taken from the data and Blue Bell returned, with the rescaling (Section 6.1). The first two steps use the 1,234 market-months in which every model's direct forecast solves, the third the 1,246 of the evaluation design, so the differences between rows are not a decomposition on identical market-months. Panel B: Blue Bell's forecast 2015 gallons with Blue Bell present and the forecast retention, per 2014 Blue Bell gallon, in the direct forecast.

The two specifications with PL-pooled nests have the lowest misplaced shares at every step: the nested logit 0.253, 0.145, and 0.101, and RCNL 0.338, 0.125, and 0.104. Every model’s misplaced share falls from the direct forecast to the evaluation design, and the specifications without nesting change order: at the direct forecast logit has the lowest of the three and the RC model the highest, and in the evaluation design the order is reversed. The direct forecasts put Blue Bell’s volume at 0.62 of its 2014 gallons for the RC model and between 0.64 and 0.84 for the other models (panel B). A direct forecast therefore ranks the models partly on their forecasts of Blue Bell’s own volume and of rivals’ 2015 sales levels; forecasting sales levels and forecasting diversion conditional on the rival market are distinct tasks, and the main comparison addresses the second. Rescaling removes the direct dependence on the forecast number of Blue Bell gallons; the forecast utility still affects which consumers choose Blue Bell and their substitution patterns. Appendix C reports the errors of the direct forecast under other rules for the residual demand and in markets without Blue Bell.

7 Rival Prices and Consumer Welfare

The models' differences in substitution carry over to the consumer value of Blue Bell's availability, and whether rival prices come from the data or from each model's pricing equilibrium changes the loss by an amount of the same order as the differences between the models. Let $\mathcal{J}$ be the choice set with Blue Bell and $\mathcal{J} \setminus \mathcal{B}$ the set without its products $\mathcal{B}$, p_0 the prices the market would have had without the recall, and p_1 the prices it had. The loss is $\Lambda = CS(\mathcal{J}, p_0) - CS(\mathcal{J} \setminus \mathcal{B}, p_1)$, where a positive value of Λ denotes a welfare loss, and it splits exactly into an availability term and a price term in either order:

$$\begin{aligned} \Lambda &= \underbrace{CS(\mathcal{J}, p_0) - CS(\mathcal{J} \setminus \mathcal{B}, p_0)}_{\text{availability at } p_0} + \underbrace{CS(\mathcal{J} \setminus \mathcal{B}, p_0) - CS(\mathcal{J} \setminus \mathcal{B}, p_1)}_{\text{rival prices without Blue Bell}} \\ &= \underbrace{CS(\mathcal{J}, p_0) - CS(\mathcal{J}, p_1)}_{\text{rival prices with Blue Bell}} + \underbrace{CS(\mathcal{J}, p_1) - CS(\mathcal{J} \setminus \mathcal{B}, p_1)}_{\text{availability at } p_1}. \end{aligned} \quad (23)$$

I use the second decomposition, so that the availability term is evaluated at the prices the market had.

Consumer surplus is measured from the inclusive value. With each consumer type's price coefficient negative and common across products, expected surplus from a logit menu is proportional to its inclusive value, and the loss per purchase opportunity from removing Blue Bell's products at given prices is

$$\ell_{mt} = \int \frac{\mathcal{I}_i(\mathcal{J}_{mt}) - \mathcal{I}_i(\mathcal{J}_{mt} \setminus \mathcal{B})}{|\alpha_i|} dF_m(i), \quad (24)$$

where $\mathcal{I}_i = \log(1 + \sum_j \exp v_{ijmt})$ for conditional logit and the nested inclusive value for nested logit, and $\alpha_i < 0$ is consumer i 's price coefficient. The measure equals the compensating variation when the marginal utility of income is constant and approximates it otherwise. In conditional logit, removing $\mathcal{B}$ lowers consumer i 's inclusive value by $-\log(1 - P_{i\mathcal{B}})$, so the loss per purchase opportunity is

$$\ell_i = \frac{-\log(1 - P_{i\mathcal{B}})}{|\alpha_i|}, \quad (25)$$

and in nested logit the inclusive value falls by $-\log \Delta_i$, with Δ_i the factor of equation (12). Because the outside option is a singleton nest with utility normalized to zero, in logit and nested logit without random tastes the inclusive value equals $-\log s_0$, and the loss in a market is the purchase opportunities times $\log(s_0^{\text{removed}}/s_0^{\text{present}})/|\alpha|$, which is positive because the outside share rises when Blue Bell is removed. Welfare is computed by consumer type where

price coefficients vary. The unit throughout is dollars per 2014 Blue Bell gallon. As in the evaluation design, the availability term is rescaled in each market–month to Blue Bell’s 2014 volume (equation (17)); the price term does not scale with Blue Bell’s volume and is summed over market–months and divided by Blue Bell’s 2014 gallons. I denote by $\tilde{\Lambda}$ the sum of the rescaled availability term and the price term: the loss of equation (23) with the availability term evaluated at Blue Bell’s 2014 volume. Table 13 reports $\tilde{\Lambda}$.

7.1 The value of Blue Bell’s availability

The PL-separate nested logit values Blue Bell’s availability about 45 percent more highly than the PL-pooled specification, despite their nearly identical price coefficients. This comparison illustrates how restrictions on substitution affect the implied value of availability (Akerberg and Rysman, 2005). Table 11 reports the availability term at p_1 , the observed 2015 rival prices. The calculation compares the observed May–August 2015 market with the same market after Blue Bell is added back, holding the fixed products at their observed quantities as in the evaluation design. The PL-separate nested logit, which keeps too little of Blue Bell’s demand among the rivals, puts the value at \$6.53 per 2014 Blue Bell gallon, and the PL-pooled nested logit, the grouping closest on both margins, at \$4.51. Blue Bell sold 5.2 million gallons in these market–months of 2014, so for May–August 2015 at the retailers the scanner data cover, the two nested logits value Blue Bell’s availability at 2015 rival prices at \$34 million and \$23 million.

Across the other models the dollar ranking need not follow the retention ranking, because the price coefficient converts the lost utility into dollars. Logit loses about twice the utility per Blue Bell gallon that the PL-pooled nested logit does, because it sends more of Blue Bell’s buyers to the outside option, but a unit of its utility is worth less than half as many dollars, so its value, \$5.25, comes close to the nested logit’s. The RC model and the group RC retain less than the nested logits yet value availability less, at \$3.30 and \$3.56, and RCNL, which retains about as much as the PL-pooled nested logit, values it at \$3.64, so the six values run from \$3.30 to \$6.53.

Table 11: The value of Blue Bell's availability at observed 2015 rival prices

	Logit	NL, PL separate	NL, PL pooled	RC	RCNL	Group RC
Retention (Table 8)	0.424	0.619	0.742	0.463	0.735	0.443
Dollars per 2014 Blue Bell gallon	5.25	6.53	4.51	3.30	3.64	3.56
Dollars per gallon of the model's Blue Bell	5.34	7.41	5.81	3.36	4.65	3.60

The availability term of equation (23) at the observed 2015 rival prices, on the market-months of the evaluation design in which the welfare calculation solves (1,244 for logit, the nested logits, and the RC model; 1,245 for RCNL and the group RC). The first dollar row rescales each market-month's term to Blue Bell's 2014 same-month gallons; the second divides it by the model's own Blue Bell gallons.

The link between retention and the availability term is exact in logit and nested logit, where the outside option is a singleton nest: for a given Blue Bell share, outside share without Blue Bell, and price coefficient, a model with higher retention implies a smaller availability term (Appendix E). For given Blue Bell and outside shares, price sensitivity, and retention, the link does not depend on how the retained demand is allocated across rivals, so different allocations can imply the same availability value, and a model's allocation error need not show up in its welfare number. The withdrawal disciplines the substitution part of these calculations but not the price coefficient that turns utility into dollars, so it does not rank the dollar values by accuracy, and the models closest on both margins do not have the most credible values for that reason alone. A removal and a price change also weight different consumers (Conlon and Mortimer, 2021), so a model that predicts a removal well need not predict merger price effects or other counterfactuals equally well.

7.2 Rival prices

Rival prices fell in Blue Bell's markets during the withdrawal. The comparison matches each product in each market with itself a year earlier, so product-market effects difference out, and compares these annual log price changes in Blue Bell's markets with those in the comparison markets, converted to percent. Relative to markets without Blue Bell, and net of the same comparison a year earlier, rival prices fell by 4.4 percent from May–August 2014 to May–August 2015, with a 95 percent interval from -7.2 to -1.5 (Table 12). Private-label prices fell by 4.6 percent and Dreyer's by 11.0. Without the adjustment for the earlier year, the decline in all rival prices is 3.1 percent. The difference in differences identifies the recall's effect on prices under the conditions of Section 3.1, that rival prices in Blue Bell's markets would otherwise have moved as in the comparison markets and that the recall left the comparison markets' prices unchanged; the adjustment subtracts the preceding year's differential change; its causal interpretation requires that this differential change would have

Table 12: Observed and predicted change in rival prices during the withdrawal, percent

	Observed		Predicted by the model's supply side					
	Net		Raw	Logit	NL, PL pooled	RC	RCNL	Group RC
<i>A. By rival group, all Blue Bell markets</i>								
All rivals	-4.40 [-7.20, -1.52]	-3.05 [-4.95, -1.12]	0.40		1.69	0.26	1.69	0.26
Private label	-4.63 [-8.58, -0.51]	-3.71 [-6.55, -0.78]	0.69		2.95	0.38	2.84	0.39
Dreyer's	-11.04 [-14.96, -6.95]	-5.48 [-8.62, -2.24]	0.18		0.40	0.16	0.71	0.20
Breyers	-3.11 [-6.07, -0.05]	0.89 [-1.40, 3.23]	0.11		0.21	0.14	0.37	0.15
Blue Bunny	-2.91 [-12.01, 7.14]	-0.50 [-4.94, 4.14]	0.09		0.39	0.11	0.51	0.12
Premium pints	-0.28 [-1.78, 1.23]	0.84 [-0.57, 2.26]	0.11		0.18	0.11	0.17	0.14
Other brands	-1.54 [-3.50, 0.46]	-2.29 [-3.59, -0.98]	0.08		0.51	0.12	0.57	0.11
<i>B. All rivals, by exposure</i>								
High	-2.66 [-6.02, 0.83]	-0.71 [-2.79, 1.42]	1.33		8.30	0.88	7.86	0.92
Medium	-6.55 [-9.19, -3.82]	-4.47 [-6.25, -2.65]	0.43		1.25	0.25	1.32	0.27
Low	-3.25 [-6.54, 0.15]	-2.59 [-4.80, -0.33]	0.18		0.58	0.12	0.63	0.11

Percent; 95 percent intervals in brackets, from 999 DMA block resamples. Observed: the change in rival prices from May–August 2014 to May–August 2015 in Blue Bell's markets minus the same change in markets that never had Blue Bell, matched within product, DMA–retailer, and month and weighted by 2014 gallons; net subtracts the same comparison for 2013 to 2014. Exposure classes split Blue Bell's markets by Blue Bell's 2014 share of DMA volume (cutoffs 7.78 and 31.27 percent). Predicted: the change in rival prices when Blue Bell is removed, with every rival repricing under the model's multiproduct Nash–Bertrand pricing (Appendix E). The PL-separate nested logit is not shown, because its recovered marginal costs are negative for about 96 percent of private-label volume.

persisted without the recall. Figure 7 shows when the decline occurred. Month by month, the net decline appears in April 2015, is largest in May and June, about 7 and 5 percent, fades through the summer, and the point estimate is close to zero by September; in 2014, the placebo year, the same comparison fluctuates around zero.

The models predict small increases. With every rival repricing under each model's multiproduct Nash–Bertrand pricing, the predicted change in rival prices when Blue Bell is removed is an increase of 0.4 percent in the exposed markets for logit, 0.3 for the RC model and the group RC, and 1.7 for the PL-pooled nested logit and RCNL, against the observed decline of 4.4 percent. The largest predicted increases are those of the PL-pooled nested logit and RCNL in the high-exposure markets, 8.3 and 7.9 percent for all rivals, where the observed net change is -2.7 percent. The decline does not increase monotonically with prior Blue Bell exposure, unlike the gradient the pricing simulations predict: it is 6.5 percent in the medium-exposure markets but 2.7 percent in the high-exposure ones. Bhattacharya et al. (2026) also find that predicted price changes in mergers are typically larger than realized ones. The two routes below therefore give the price term opposite signs.

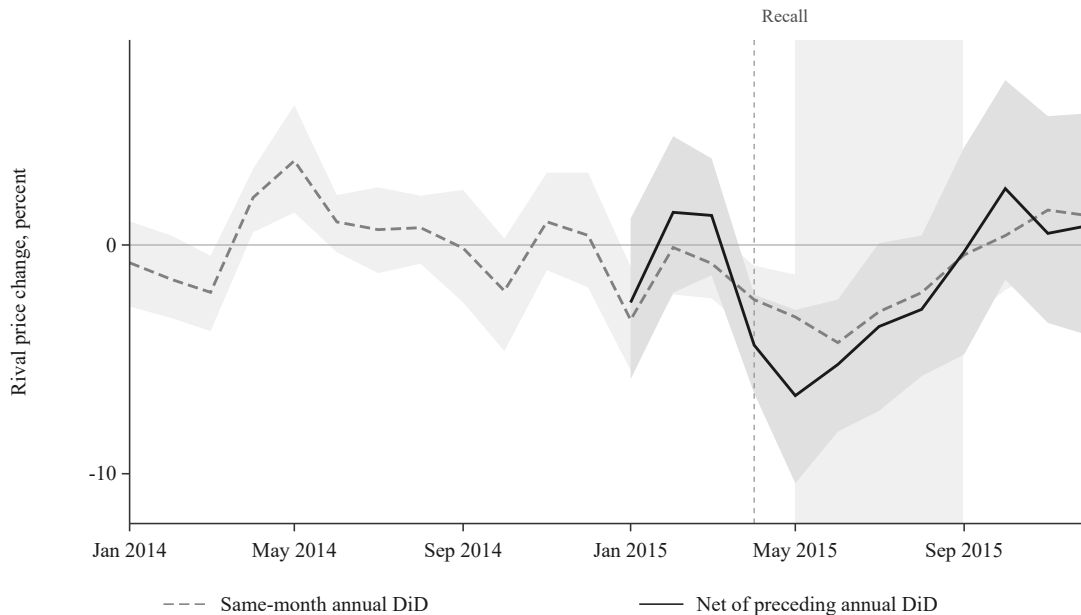


Figure 7: Rival price changes in Blue Bell's markets by month

Percent. Dashed: same-month annual difference in differences of matched rival prices, 2013 to 2014 for the 2014 months and 2014 to 2015 for the 2015 months; solid: the 2015 difference net of the preceding year's. The monthly series uses the scanner history of the counterfactual markets, whose support differs slightly from Table 12. Bands: pointwise 95 percent intervals from 999 DMA block resamples.

7.3 The loss with price responses

Table 13 reports $\tilde{\Lambda}$ under two routes for the prices. Under the observed-price route, p_1 is the observed 2015 price, as it was, which may include responses to the recall and other changes, and p_0 takes each rival's price back by the net observed change of Table 12, by rival group and exposure class. Under the model-price route, p_0 and p_1 are each model's multiproduct Nash–Bertrand prices with and without Blue Bell, given recorded ownership, retailers pricing their own private label, and marginal costs recovered before the withdrawal (Appendix E); the route adds the assumptions of the supply model, and I do not compute it for the PL-separate nested logit, whose recovered marginal costs are negative for about 96 percent of private-label volume, against at most about a fifth in the other models. Under model prices the price term is the exact value of the price change at the full choice set, and under observed prices its first-order value; in both routes the availability term is evaluated at p_1 , and the loss is their sum.

Under observed prices the price term is a gain, because rival prices fell. I value it at first order, the sum over rivals of observed quantity times the observed price change, \$2.28 per 2014 Blue Bell gallon in every model. Because consumer surplus is convex in prices, and

Table 13: The consumer loss from the withdrawal with price responses

	Observed prices			Model prices		
	Availability at p_1	Price term	Lower bound on $\tilde{\Lambda}$	Availability at p_1	Price term	Model-implied $\tilde{\Lambda}$
Logit	5.25	-2.28	2.98	5.35	0.18	5.52
NL, PL separate	6.53	-2.28	4.26	n.r.	n.r.	n.r.
NL, PL pooled	4.51	-2.28	2.24	4.17	0.88	5.05
RC	3.30	-2.28	1.02	3.61	0.13	3.74
RCNL	3.64	-2.28	1.37	4.14	0.82	4.96
Group RC	3.56	-2.28	1.29	3.67	0.13	3.80

Dollars per 2014 Blue Bell gallon; a positive value is a loss. The availability term is rescaled as in Table 11 and evaluated at each route's p_1 . Observed prices: p_1 observed and p_0 backed out from the net observed change of Table 12; the price term is the first-order value of that change, the same for every model. Model prices: p_0 and p_1 from each model's multiproduct Nash–Bertrand pricing (Appendix E), on the market-months in which its price equilibrium solves (1,234 for logit, the PL-pooled nested logit, and the RC model; 1,235 for RCNL; 1,229 for the group RC). n.r.: not reported, because the model's recovered marginal costs are negative for about 96 percent of private-label volume.

rival quantities with Blue Bell present are below those observed, this value bounds the gain from above when rival prices fall, and the loss it implies is a lower bound: \$1.02 to \$4.26 per 2014 Blue Bell gallon across the six models. Under model prices the price term adds \$0.13 to \$0.88 to the loss, so the loss is \$3.74 to \$5.52. The price route thus changes the loss by an amount of the same order as the differences between the models' availability values. So the fixed-price comparison of substitution does not by itself settle the welfare loss once rival prices respond.

8 Sensitivity

The retention comparison is sensitive to the benchmark and the assumed market size. The adjustments below narrow the differences across models and, when sufficiently large, change their ranking. Under each change other than the larger adjustments for the rival price response, one of the two specifications with PL-pooled nests keeps the lowest misplaced share.

Subtracting the pre-recall differential trend lowers benchmark retention to 0.65, while private label still accounts for about half of the rival gain. The adjustment uses the same difference in differences applied to 2014 against 2013, 0.12 per Blue Bell gallon (Table 3). The PL-separate nested logit, with retention of 0.62, then comes closer on retention than the PL-pooled nested logit and RCNL, at 0.74 and 0.73. RCNL nevertheless has the lowest misplaced share, 0.15, followed by the PL-pooled nested logit at 0.18 and the RC model at

0.22. The two models with PL-pooled nests retain an advantage on the combined score, but the difference from the non-nested models is smaller.

Rival sales induced by the price response to the recall may also be part of the observed gain, although the models are evaluated at fixed prices. To examine how much this matters, I subtract a hypothesized price-induced gain in proportion to each destination’s observed gain. This adjustment lowers benchmark retention while holding allocation shares fixed. A model with PL-pooled nests has the lowest misplaced share until benchmark retention falls to about 0.59: the nested logit above 0.77 and RCNL between 0.77 and 0.59. Larger reductions favor the RC model and, below about 0.45, logit (Appendix F). Thus, attributing a sufficiently large part of the rival gain to lower prices can reverse the main ranking.

Excluding private label from the scored destinations narrows the difference between the two nested logits. RCNL has the lowest misplaced share among the five branded destinations, 0.07, and a lower score than the PL-pooled nested logit in 82 percent of paired draws. The PL-pooled nested logit remains ahead of the PL-separate one in 70 percent of draws, but their point estimates are close. Much of the difference between the two nested logits therefore concerns their predictions for private label.⁸

Reducing the assumed market size by 20 percent primarily changes predicted retention (Table 14). With parameters held fixed, retention rises by about 0.10 in logit, the RC model, and the group RC, compared with 0.02 to 0.05 in the nested specifications; allocation errors change by at most 0.04. The nested specifications are therefore less sensitive to this market-size change.⁹ Lower sensitivity to the market size is not evidence of greater accuracy. RCNL has the lowest misplaced share, 0.08, followed by the PL-pooled nested logit at 0.09. The RC model ranks third, and the scores of all three non-nested specifications fall by 0.08 to 0.09. As with a lower retention benchmark, a smaller market size narrows the non-nested models’ retention shortfall.

For the separate direct-forecast exercise, I also vary the pre-recall window used to forecast each product’s demand residual. Across four alternative windows, including one that excludes March 2015, misplaced shares change by up to 0.07. The PL-pooled nested logit has the lowest score under every window among the five models evaluated in this check (Appendix C).

⁸The PL-pooled nested logit’s score is 0.101 with or without private label because changes in the destination-error and retention-error components offset each other.

⁹Miller and Weinberg (2017), Supplemental Material, Section F, instead re-estimate beer demand at market sizes of 1.3 and 2 times maximum observed sales. The nesting parameter adjusts, and median diversion to the outside good stays between 13 and 16 percent. Here the parameters remain fixed.

Table 14: Predictions at a smaller market size

	Main market size			0.8 times the main size		
	Retention	Allocation error	Misplaced share	Retention	Allocation error	Misplaced share
Observed	0.774	.	.	0.774	.	.
Logit	0.424	0.106	0.350	0.532	0.107	0.259
NL, PL separate	0.619	0.329	0.291	0.669	0.294	0.257
NL, PL pooled	0.742	0.118	0.101	0.763	0.109	0.090
RC	0.463	0.131	0.322	0.562	0.128	0.245
RCNL	0.735	0.091	0.104	0.770	0.093	0.077
Group RC	0.443	0.134	0.340	0.547	0.130	0.259

Evaluation design of Section 6.1 on the 1,246 market-months, with every model's parameters held at their estimates. The main market size is 1.5 times each DMA-retailer's largest monthly inside volume through March 2015, and the smaller one 1.2 times that volume. Retention in gallons gained by rivals per 2014 Blue Bell gallon.

9 Conclusion

Rivals recaptured about three quarters of Blue Bell's former volume during the 2015 withdrawal, but the models differed in both retention and the allocation of that demand. Under the main benchmark, the nested logit that places private label with Blue Bell's mainstream tubs and RCNL on the same nests combine retention close to the benchmark with relatively small allocation errors. Logit and the random-coefficient specifications without nesting retain too little demand. Giving private label a separate nest also reduces retention and sends too little of the rival gain to private label. The differences narrow under the alternative retention benchmarks, and sufficiently large adjustments for the rival price response can change the ranking; under the other checks of Section 8, a specification with PL-pooled nests keeps the lowest misplaced share.

The comparison favors specifications that represent strong substitution between Blue Bell and private label. RCNL has a slightly smaller allocation error than the PL-pooled nested logit, but their retention predictions and misplaced shares are close. The group RC uses the same product groups, yet its estimated group tastes do not close the retention gap. The distinction is therefore between how the fitted models represent substitution, not simply whether they include random coefficients. Allowing additional taste heterogeneity yields little improvement over the PL-pooled nested logit in the main comparison, so the grouping deserves to be reported and varied like any other specification choice.

The grouping also affects the implied value of availability. With nearly identical price coefficients, the PL-separate nested logit values Blue Bell's availability about 45 percent more highly than the PL-pooled nested logit; across the six models, the value ranges from

\$3.3 to \$6.5 per 2014 Blue Bell gallon. Rival prices add a separate source of variation in welfare: the models predict price increases, whereas observed rival prices fell by 4.4 percent relative to markets without Blue Bell. Using observed rather than model-predicted prices changes the implied loss by an amount comparable to the differences in availability values across models.

Evaluating retention and allocation separately reveals different errors: a model can retain much of the removed demand but send it to the wrong rivals, or divide retained demand accurately while losing too much to the outside option. The comparison evaluates substitution at observed rival prices and sales; it does not by itself validate the price responses used in welfare or merger simulations.

References

- Akerberg, Daniel A. and Marc Rysman (2005), “Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects.” *RAND Journal of Economics*, 36, 771–788.
- Berry, Steven, James Levinsohn, and Ariel Pakes (1995), “Automobile Prices in Market Equilibrium.” *Econometrica*, 63, 841–890.
- Berry, Steven, James Levinsohn, and Ariel Pakes (2004), “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market.” *Journal of Political Economy*, 112, 68–105.
- Berry, Steven T. (1994), “Estimating Discrete-Choice Models of Product Differentiation.” *RAND Journal of Economics*, 25, 242–262.
- Berry, Steven T. and Philip A. Haile (2021), “Foundations of Demand Estimation.” Cowles Foundation Discussion Paper, Cowles Foundation for Research in Economics, Yale University.
- Bhattacharya, Vivek, Gastón Illanes, Avner A. Kreps, José D. Salas, and David Stillerman (2026), “A Large-Scale Evaluation of Merger Simulations.” NBER Working Paper, National Bureau of Economic Research.
- Björnerstedt, Jonas and Frank Verboven (2016), “Does Merger Simulation Work? Evidence from the Swedish Analgesics Market.” *American Economic Journal: Applied Economics*, 8, 125–164.

- Cardell, N. Scott (1997), “Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity.” *Econometric Theory*, 13, 185–213.
- Centers for Disease Control and Prevention (2015), “2015 Outbreak of Listeria Infections Linked to Blue Bell Creameries Products.” https://archive.cdc.gov/www_cdc_gov/listeria/outbreaks/ice-cream-03-15/index.html, June 10.
- Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi (2003), “Why Don’t Prices Rise During Periods of Peak Demand? Evidence from Scanner Data.” *American Economic Review*, 93, 15–37.
- Cleeren, Kathleen, Harald J. van Heerde, and Marnik G. Dekimpe (2013), “Rising from the Ashes: How Brands and Categories Can Overcome Product-Harm Crises.” *Journal of Marketing*, 77, 58–77.
- Conlon, Christopher and Jeff Gortmaker (2020), “Best Practices for Differentiated Products Demand Estimation with PyBLP.” *RAND Journal of Economics*, 51, 1108–1161.
- Conlon, Christopher and Jeff Gortmaker (2025), “Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP.” *Journal of Econometrics*, 105926.
- Conlon, Christopher and Julie Holland Mortimer (2021), “Empirical Properties of Diversion Ratios.” *RAND Journal of Economics*, 52, 693–726.
- Conlon, Christopher T. and Julie Holland Mortimer (2013), “Demand Estimation Under Incomplete Product Availability.” *American Economic Journal: Microeconomics*, 5, 1–30.
- Draganska, Michaela, Michael Mazzeo, and Katja Seim (2009), “Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing Decisions.” *Quantitative Marketing and Economics*, 7, 105–146.
- Fan, Ying and Chenyu Yang (2025), “Estimating Discrete Games with Many Firms and Many Decisions: An Application to Merger and Product Variety.” *Journal of Political Economy*, 133, 1886–1931.
- Federal Trade Commission (2003), “Complaint, In the Matter of Nestlé Holdings, Inc., Dreyer’s Grand Ice Cream Holdings, Inc., and Dreyer’s Grand Ice Cream, Inc.” Docket No. C-4082, <https://www.ftc.gov/sites/default/files/documents/cases/2003/06/dreyercomplaint.htm>, June 25.

- Freedman, Seth, Melissa Kearney, and Mara Lederman (2012), “Product Recalls, Imperfect Information, and Spillover Effects: Lessons from the Consumer Response to the 2007 Toy Recalls.” *Review of Economics and Statistics*, 94, 499–516.
- Galichon, Alfred (2022), “On the Representation of the Nested Logit Model.” *Econometric Theory*, 38, 370–380.
- Gandhi, Amit and Jean-François Houde (2020), “Measuring Substitution Patterns in Differentiated-Products Industries.” Working paper.
- Grigolon, Laura and Frank Verboven (2014), “Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation.” *Review of Economics and Statistics*, 96, 916–935.
- Kennedy, Sarah M. (2013), “Ice Cream Sales Are Up, Novelties Are Melting.” *Dairy Foods*, <https://www.dairyfoods.com/articles/89756-ice-cream-sales-are-up-novelties-are-melting>, IRI data, 52 weeks ended May 19, 2013.
- Miller, Nathan H. and Matthew C. Weinberg (2017), “Understanding the Price Effects of the MillerCoors Joint Venture.” *Econometrica*, 85, 1763–1791.
- Pathak, Parag A. and Peng Shi (2021), “How Well Do Structural Demand Models Work? Counterfactual Predictions in School Choice.” *Journal of Econometrics*, 222, 161–195.
- Peters, Craig (2006), “Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry.” *Journal of Law and Economics*, 49, 627–649.
- Petrin, Amil (2002), “Quantifying the Benefits of New Products: The Case of the Minivan.” *Journal of Political Economy*, 110, 705–729.
- Raval, Devesh, Ted Rosenbaum, and Nathan E. Wilson (2022), “Using Disaster-Induced Closures to Evaluate Discrete Choice Models of Hospital Demand.” *RAND Journal of Economics*, 53, 561–589.
- Sayman, Serdar, Stephen J. Hoch, and Jagmohan S. Raju (2002), “Positioning of Store Brands.” *Marketing Science*, 21, 378–397.
- Sullivan, Christopher (2020), “The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion.” Working paper, University of Wisconsin-Madison.

- Thomsen, Michael R., Rimma Shiptsova, and Sandra J. Hamm (2006), “Sales Responses to Recalls for *Listeria monocytogenes*: Evidence from Branded Ready-to-Eat Meats.” *Review of Agricultural Economics*, 28, 482–493.
- Train, Kenneth E. (2009), *Discrete Choice Methods with Simulation*. 2nd edition, Cambridge: Cambridge University Press.
- U.S. Department of Agriculture, Agricultural Marketing Service (2003), “Commercial Item Description: Ice Cream, Sherbet, Fruit and Juice Products, Ices and Novelties.” Commercial Item Description, USDA Agricultural Marketing Service, <https://www.ams.usda.gov/sites/default/files/media/CID%20Ice%20Cream%2C%20Sherbet%2C%20Fruit%20and%20Juice%20Bars%2C%20Ices%2C%20and%20Novelties.pdf>, Quality segments defined in Section 5.1.
- U.S. Food and Drug Administration (2015), “Blue Bell Creameries: Inspection Records (FOIA Electronic Reading Room).” <https://www.fda.gov/about-fda/oii-foia-electronic-reading-room/blue-bell-creameries>, Form FDA 483 issued for Brenham, TX (May 1, 2015), Broken Arrow, OK (April 23, 2015), and Sylacauga, AL (April 30, 2015).
- Velasco, Schuyler (2015), “Blue Bell Ice Cream Recalls All Products amid *Listeria* Outbreak.” *The Christian Science Monitor*, <https://www.csmonitor.com/Business/The-Bite/2015/0421/Blue-Bell-Ice-Cream-recalls-all-products-amid-listeria-outbreak>, April 21.
- Verboven, Frank (1996), “International Price Discrimination in the European Car Market.” *RAND Journal of Economics*, 27, 240–268.
- Wallace, Terry and Nicholas Clayton (2015), “Blue Bell Recall Focused on Ice Cream Bars, Not Half Gallons, Quarts, or Pints.” Associated Press, via *The Christian Science Monitor*, <https://www.csmonitor.com/Business/2015/0314/Blue-Bell-recall-focused-on-ice-cream-bars-not-half-gallons-quarts-or-pints>, March 14.
- Weinberg, Matthew C. and Daniel Hosken (2013), “Evidence on the Accuracy of Merger Simulations.” *Review of Economics and Statistics*, 95, 1584–1600.
- Zhang, Linqi (2026), “Identification and Estimation of Market Size in Discrete Choice Demand Models.” https://zhanglinqi.github.io/assets/pdf/market_size_Linqi_Zhang.pdf, Working paper, The Chinese University of Hong Kong Business School.

A Data Construction and Product Definitions

A.1 Units, aggregation, and availability

The scanner observations begin at the product, store, and period level, and package size converts packages into gallons. Price is expenditure divided by gallons, computed after summing expenditure and gallons over the observations that are combined. Averaging package prices or unweighted unit values instead would give small and large packages different weights and would create price changes whenever the package mix changes.

An estimated product combines a brand with the characteristics that enter utility. The bulk, bar, cone, and sandwich indicators come from Nielsen’s form code through a crosswalk, and items without a recorded form are treated as bulk ice cream.

An observed sale shows that a product was available, while an absence of sales can reflect no purchases, a missing store observation, or no availability. Monthly aggregation cannot resolve short stockouts, so the estimation menu is the set of products with recorded sales in the market and month. Blue Bell’s absence is set by the recall; the rivals’ assortments are set by recorded sales.

The market is a DMA paired with a retailer, because retailer identity matters both for the choice set and for private label. The product effect absorbs time-invariant appeal, the DMA–retailer effect absorbs local differences, and the retailer-by-private-label effects in the main specification let private label differ in mean appeal across retailers.

Table A.1 lists the counts behind estimation and the forecasts. Two cases separate the two: a product estimated but without the price history the forecast needs, which is held at its observed gallons, and a product sold in the withdrawal period but not among the estimated products, which is held at its observed gallons and takes the group of its own first-quarter 2015 sales (Appendix C).

A.2 Private label, the sample by year, and exposure

Table A.2 lists the four private-label products. Each is one masked Nielsen private-label brand code with no characteristic information, so its characteristic indicators are zero and its random tastes in the RC model come only from the inside-good taste and price. A market is one DMA–retailer pair, so the private-label product in a market is that retailer’s store brand, and the retailer-chain-by-private-label effects of the main specification give each chain’s private label its own mean utility. “CTL BR - IC” carries 205.2 of the 276.1 million private-label gallons in 2011–2014.

Table A.3 reports the estimation sample by year. Each full year contributes between 0.84

Table A.1: Product and market coverage for estimation and forecasting

Object	Count or volume	Role
Estimation products	756	Fixed product identifiers
DMA-retailer markets	529	Estimation markets
DMAs	117	
DMAs with Blue Bell sales in 2014, estimation sample	105	Summary statistics
DMAs with positive Blue Bell scanner sales, forecast definition	103	Exposed markets
Scanner-zero DMAs	14	Category-volume denominator
Blue Bell 2014 volume share, 105 DMAs with sales in the estimation sample	12.8% (7.8% in all DMAs)	Scale of the removal
Scored product-market-months, 1,246 market-months	155,101	Scored withdrawal sample
Observed rival gallons in scored products, Blue Bell markets	29.8 million	Scored volume
Observed rival gallons held fixed, Blue Bell markets	0.66 million	Conditioning input
Local residual: January–March 2015, earlier local months, or zero	97.3, 0.7, and 1.8% of rival gallons	Table C.1
Group fallback (own first-quarter 2015 sales)	5.9% of non-estimation gallons; 0.26% of rival gallons in Blue Bell markets	Group assignment

Counts refer to the estimation files. Forecast counts and volumes refer to May–August 2015: the scored product-market-months are those of Table 1, and the gallons and residual shares are rival gallons in Blue Bell’s markets under the main RC’s forecast rule (Table C.1), which every specification shares.

Table A.2: Private-label products

Product	Nielsen brand code	Nielsen label	DMAs	DMA-retailer markets	Chains	Gallons, 2011–2014 (millions)
473	536746	CTL BR	117	487	53	53.9
474	536757	CTL BR - IC	117	484	50	205.2
475	536758	CTL BR - IM	61	68	9	0.8
476	536764	CTL BR - SH	117	476	49	16.2
All private label						276.1

Nielsen masks private-label items: each carries one of the four generic private-label brand codes and no characteristic information, so each code is one model product with every characteristic indicator equal to zero. Chains count the retailer chains with sales of the product. Gallons follow the volume convention of Table 2.

and 0.91 million observations from 655 to 717 model products, and the first quarter of 2015 adds 0.20 million observations from 650 products; the sample has 655 products in 2011, 688 in 2012, 717 in 2013, and 702 in 2014, out of 756 overall, so some products enter and others leave before April 2015. Prices are stable across years, with row averages between \$28.29 and \$29.05 per gallon and volume-weighted averages between \$13.37 and \$13.84. Bulk containers are 46 percent of products and 43 percent of volume, while chocolate products are 39 percent of products but 21 percent of volume, vanilla 30 and 18 percent, low fat 20 and 11 percent, low calorie 23 and 12 percent, and cookies or mix-ins 10 and 4 percent.

Table A.3: Estimation sample by year and by characteristic

<i>Panel A. Panel structure</i>						
	2011	2012	2013	2014	2015 Q1	All
Product–market–months	841,530	906,492	896,472	858,389	203,075	3,705,958
Model products	655	688	717	702	650	756
DMA–retailer markets	529	529	529	529	529	529
DMAs	117	117	117	117	117	117
Months	12	12	12	12	3	51
Volume (million gallons)	172.7	173.7	171.4	170.7	35.6	724.0
<i>Panel B. Price per gallon (\$), product–market–months</i>						
	2011	2012	2013	2014	2015 Q1	All
Mean	28.39	29.00	28.51	28.29	29.05	28.58
Standard deviation	16.30	16.41	16.72	16.88	17.65	16.64
10th percentile	10.29	10.20	9.73	9.58	10.08	9.94
Median	26.96	27.77	26.73	25.97	25.90	26.84
90th percentile	48.67	49.81	50.30	50.56	53.51	50.08
Volume-weighted mean	13.45	13.57	13.37	13.42	13.84	13.47
<i>Panel C. Characteristics, all years (shares)</i>						
	Products	Rows	Volume			
Vanilla	0.302	0.296	0.176			
Chocolate	0.388	0.412	0.208			
Bulk	0.463	0.397	0.433			
Low fat	0.202	0.186	0.106			
Low calorie	0.228	0.251	0.125			
Assorted	0.165	0.163	0.134			
Cookies or mix-ins	0.103	0.115	0.041			

Estimation file for all specifications, January 2011 to March 2015. Model products aggregate items that differ only in package size. Volume is the inside share times the market size, in gallons; prices are expenditure per gallon in 2019 dollars. Panel C reports the share of model products, of product–market–month rows, and of volume with each indicator.

Two exposure definitions appear in the paper. The estimation sample records Blue Bell sales in 105 DMAs in 2014, and Table 2 uses that definition. The counterfactuals cover a wider set of DMA–retailer markets and classify them by Blue Bell’s scanner sales. The scored market–months are those with Blue Bell sales in the same month of 2014, in 103 DMAs, and the comparison markets of the observed effect are the 266 DMA–retailer markets, in 65

DMAs, with no Blue Bell scanner sales in any month from 2011 through 2015. Markets in three DMAs with Blue Bell sales in 2011–2014 but none in May–August 2014 or in 2015, and in three DMAs with a few recorded Blue Bell sales in 2015, 4 gallons in total, belong to neither group. The comparison DMAs were also required to show no Blue Bell purchases by panel households, and DMAs with zero scanner sales but household purchases of Blue Bell are excluded from both groups. Neither screen shows that no household in a comparison DMA could buy the brand, since households can shop outside covered stores; if some could, the comparison markets absorb a little of the withdrawal and the observed effect understates it.

A.3 Market size and product groups

The market size of DMA–retailer m is

$$M_m = 1.5 \max_{t \leq \text{Mar 2015}} Q_{mt}^{inside}, \quad (\text{A.1})$$

the largest monthly inside volume over the months through March 2015, held fixed over time; in the full-window estimates the maximum runs through 2016. The definition uses only volume from the estimation window in the main estimates and is the same for every specification. It measures purchase opportunities in gallons.

Both groupings split tubs into mainstream and premium by their pre-withdrawal brand–format price, with a threshold of \$20 per gallon, and novelties at \$25 per gallon; the PL–separate grouping gives private label a group of its own, and the PL–pooled grouping places it among the mainstream tubs (Section 4). The variant with novelties in one group (Appendix Table B.1) does not split them by tier. Because the groups are defined from pre-withdrawal prices, withdrawal-period price changes do not move a product between groups.

Brands without a pre-withdrawal group assignment take the group of their own first-quarter 2015 sales. This fallback decides 5.9 percent of the gallons outside the estimation products and 0.26 percent of rival gallons in Blue Bell markets. The same rule applies to every specification, and no group is defined by reference to a brand’s post-recall performance.

A.4 Household coverage

The household panel records purchases rather than all opportunities to buy. A household with no recorded ice cream purchase in a window may still buy ice cream elsewhere, and a missed trip can change an annual buying indicator. Roster balancing holds the set of reporting households fixed across the comparison windows, while differences in product cov-

erage between the household panel and the scanner stores remain and are reported with each statistic.

The household targets and the recall benchmark are separate calculations. The targets use 2011–2014 households and a covariance from 500 whole-household bootstrap resamples; the benchmark follows a cohort of households from 2013 through 2015 with its own whole-household bootstrap.

B Estimation Details

B.1 Parameters and the income coordinate

Table B.1 defines the variants of the specifications in Table 6; Table 9 scores them.

Table B.1: Variants of the specifications

Name	Substitution structure	Retailer $\times$ PL effects	Micro moments
NL, PL separate, novelties in one nest	Common ρ within nests; one novelty nest	Yes	None
RC without retailer $\times$ PL effects	As RC	No	Base
RC, demographic	As RC	Yes	Base and income-group contrasts

Names, moments, and instruments as in Table 6; the income-group contrasts are eight moments (Section 4.3). The first variant’s groups are mainstream tubs, premium tubs, novelties, and private label. Appendix Tables B.2 and B.3 report their estimates and Table 9 their withdrawal scores.

The nonlinear parameters of the RC model are the standard deviations of the tastes for vanilla, chocolate, bulk, low fat, low calorie, assorted products, and cookies or mix-ins, the standard deviation of the inside-good taste, the price–income interaction γ_p , and the two income-group shifts λ_{mid} and λ_{high} . RCNL adds ρ , and the group RC adds four group standard deviations. The optimizer works with the logarithms of the standard deviations; the tables report standard deviations on their level scale. For each value of the nonlinear parameters, the mean utilities that match observed shares are recovered, the product, DMA–retailer, and year effects (and the retailer-by-private-label effects where present) are absorbed, and the month-of-year effects and α are concentrated out, twelve linear parameters in all.

Income enters in a single coordinate. For a simulated consumer in DMA d , $l_i = \bar{l}_d + s_d z_i$, where $\bar{l}_d$ and s_d are the mean and standard deviation of log household income in the DMA and z_i is standard normal. Each DMA has one fixed set of 100 draws, shared by all of its retailers and used in every month. Consumer i ’s price coefficient is $\alpha + \gamma_p l_i$, so α is its intercept and γ_p its change per unit of log income. The income shifts use the same draws: a simulated consumer belongs to the low, middle, or high group according to the two log-income cutoffs that define the household income rows.

The DMA income distribution comes from the household income recorded in the Home-scan panelist file. Income brackets are mapped to dollar midpoints, with the top bracket of the 2011–2014 codebook, \$100,000 and over, at \$150,000, and then logged; records with missing income are dropped. Within each DMA, $\bar{l}_d$ and s_d are the mean and standard devi-

ation of log income weighted by the panel projection factor, computed over all panel years in the panelist extract, and DMAs with fewer than five panelist records take the unweighted national mean and standard deviation. The income groups are cut at \$50,000 and \$100,000.

Table B.2 reports the nested logit with novelties in one nest, and Table B.3 the RC specifications outside the main comparison: the RC without retailer-by-private-label effects, the demographic extension, and the estimates on the full 2011–2016 window.

Table B.2: Demand estimates: the nested logit with novelties in one nest

NL, PL separate, novelties in one nest	
Price coefficient, α	−0.096 (0.005)
Nesting parameter, ρ	0.835 (0.010)
Own-price elasticity, gallon-weighted median	−4.66
Retailer $\times$ private label	Yes
Instruments	Base and within-nest
Product–market–months	3,705,958

Definitions as in Table 7. Nests: mainstream tubs, premium tubs, novelties, and private label (Appendix Table B.1). Standard errors clustered by DMA; the Sanderson–Windmeijer first-stage F statistics for price and for the within-nest share are 332 and 377.

B.2 The instruments of RCNL

RCNL uses the within-nest instruments and the characteristic-sum and rival-count instruments but not the sugar-price instruments. Table B.4 reports RCNL under three choices of instruments and micro moments. With the sugar-price instruments, its gallon-weighted median own-price elasticity is -7.7 and its direct forecast of Blue Bell’s 2015 volume is 0.38 of the 2014 level, both outside the range of the other specifications, -1.9 to -4.5 and 0.62 to 0.84. Without the within-nest instruments and with the four own-group moments also targeted, the elasticity is -16.5 and the forecast volume 0.10. Without the sugar-price instruments, the elasticity, -4.2 , and the forecast volume, 0.64, lie within those ranges, and this is the RCNL of the comparison.

B.3 Nest size and retention in nested logit

Equation (12) implies that, for a given Blue Bell share and ρ , a larger nest around Blue Bell raises retention. Write $x = s_{\mathcal{B}|g}$ for Blue Bell’s share of its nest and $s_{\mathcal{B}} = s_g x$ for its share of the market. The nest’s term in $1 - \Delta$ is $s_g [1 - (1 - x)^{1-\rho}] = s_{\mathcal{B}} h(x)/x$ with $h(x) = 1 - (1 - x)^{1-\rho}$. Because h is convex with $h(0) = 0$, $h(x)/x$ rises with x . Holding $s_{\mathcal{B}}$ fixed, a larger nest means a smaller x , a smaller $1 - \Delta$, and so a smaller gain $s_0(1/\Delta - 1)$ for the outside option.

Table B.3: Demand estimates: other random-coefficient specifications

	Data through March 2015		Full 2011–2016 window		
	(1)	(2)	(3)	(4)	(5)
	RC, no retailer $\times$ PL effects	RC, demographic	RC RC	RC, demographic	RC, no retailer $\times$ PL effects
Price coefficient, α	−0.478 (0.024)	−0.479 (0.021)	−0.503 (0.030)	−0.500 (0.022)	−0.507 (0.031)
Price $\times$ log income, γ_p	0.0126 (0.0013)	0.0120 (0.0007)	0.0122 (0.0019)	0.0116 (0.0006)	0.0122 (0.0021)
Inside-good SD, σ_0	0.917 (0.074)	0.985 (0.066)	1.069 (0.068)	1.081 (0.054)	1.062 (0.071)
Vanilla SD	0.657 (0.034)	0.683 (0.033)	0.620 (0.036)	0.594 (0.034)	0.621 (0.036)
Chocolate SD	0.775 (0.030)	0.767 (0.030)	0.727 (0.026)	0.733 (0.025)	0.726 (0.026)
Bulk SD	0.911 (0.027)	0.908 (0.026)	0.834 (0.025)	0.835 (0.023)	0.834 (0.025)
Low-fat SD	0.850 (0.078)	0.770 (0.064)	0.999 (0.126)	0.847 (0.059)	0.998 (0.136)
Low-calorie SD	0.830 (0.049)	0.895 (0.042)	1.011 (0.043)	1.064 (0.035)	1.010 (0.044)
Assorted SD	0.022 (1.241)	0.222 (0.148)	0.021 (0.991)	0.167 (0.156)	0.018 (1.129)
Cookies / mix-ins SD	0.114 (1.853)	0.058 (1.330)	0.084 (6.028)	0.051 (1.114)	0.085 (6.337)
Middle income, λ_{mid}	0.028 (0.042)	0.030 (0.041)	0.059 (0.052)	0.048 (0.037)	0.057 (0.054)
High income, λ_{high}	−0.559 (0.052)	−0.573 (0.053)	−0.578 (0.055)	−0.563 (0.045)	−0.577 (0.056)
Product–market–months	3,705,958	3,705,958	5,080,142	5,080,142	5,080,142
Retailer $\times$ private-label effects	No	Yes	Yes	Yes	No
Instruments	Base	Base	Base	Base	Base
Micro moments	Base	Base and contrasts	Base	Base and contrasts	Base

Definitions follow Table 7; every column includes product, DMA–retailer, year, and month-of-year fixed effects. Column (1) re-estimates the main RC without the retailer-by-private-label effects; the withdrawal comparison instead evaluates the RC without these effects at the main RC’s nonlinear parameters with year effects in their place. Column (2) adds eight income-group contrast moments to the main RC. Columns (3) to (5) are estimated on 2011–2016 with the same products and markets, so they use the withdrawal and the return. The demographic columns use more households because their income-share and paid-price moments also include purchases below the base block’s annual quantity threshold (86,307 households and 221,658 household-years in column (2)); the other full-window columns use micro moments from 2011–2016 (102,483 households and 323,138 household-years in the base block).

Table B.4: RCNL under alternative instruments and micro moments

	Within-nest, no sugar prices	Within-nest and sugar prices	Base, no within-nest
Micro moments	Base	Base	Base and own-group
Price coefficient, α	-0.233	-0.369	-0.326
Nesting parameter, ρ	0.718	0.706	0.868
Own-price elasticity, gallon-weighted median	-4.20	-7.69	-16.55
Direct forecast of Blue Bell's volume, share of 2014 gallons	0.64	0.38	0.10

All three on the PL-pooled nests, without retailer-by-private-label effects, on data through March 2015. The first column is the RCNL of the comparison (Table 7). Base instruments: characteristic sums, rival counts, and the sugar price with its interactions; within-nest instruments as in Section 4.3. The other five specifications have elasticities from -1.9 to -4.5 and forecast Blue Bell volumes from 0.62 to 0.84.

B.4 The income terms

The income level enters each RC specification through the two income-group shifts, and Table B.5 tests this choice. Each test compares a restricted specification with one that contains both the two shifts and a log income level, under the same weights. Let Q_U be the minimized objective of the unrestricted specification and Q_R that of a restricted one; under the restriction, $Q_R - Q_U$ is asymptotically χ^2 with as many degrees of freedom as restrictions.

The main RC rejects a log-linear income level in favor of the group shifts. Dropping the two shifts while keeping a log income level raises the objective at the best points found by 39.8, against a 5 percent critical value of 5.99 with two restrictions, and dropping both raises it by 142.1, against 7.81 with three. Neither point is shown to be a global minimum, so these are differences at the points found rather than bounds on the distance statistic. With the shifts included, the specification without a log income level attains the lowest objective found, 484.55 against 485.66 for the best point found with the level. The RC without retailer-by-private-label effects and the two full-window RC estimates show the same pattern (Panel B). The shifts match the household income moments, which are defined on the same three groups, whereas a log income level adds variation within groups that those moments do not measure.

In the demographic extension the log income level adds information beyond the shifts. Dropping it raises the objective at the best points found by 10.6, and by 14.9 in the full-window version, each against a critical value of 3.84, so the shifts do not fully replace the level in these two specifications. Their unrestricted estimate of the log income coefficient lies on the boundary of the parameter space, so the χ^2 critical value is an approximation. RCNL uses the two group shifts for comparability, so the income terms are specified consistently across the RC models.

Table B.5: Distance-metric tests of the income terms

<i>Panel A. Restrictions that raise the objective</i>				
Specification	Restriction	Restrictions	Statistic	5% critical value
RC (main)	Drop the shifts, keep a log income level	2	39.8	5.99
RC (main)	Drop both	3	142.1	7.81
RC, demographic	Drop the log income level	1	10.6	3.84
Full window demographic	Drop the log income level	1	14.9	3.84
<i>Panel B. Objective without the log income level and lowest objective found with it</i>				
Specification	Without		With	
RC (main)	484.55		485.66	
RC without retailer $\times$ PL effects	490.33		511.16	
Full window RC (main)	593.11		594.22	
Full window RC without retailer $\times$ PL effects	594.33		595.05	

The unrestricted specification contains both the two income-group shifts and a log income level; objectives are evaluated under its weights. Panel A reports the increase in the objective at the best points found when the restriction is imposed; in the main RC the unrestricted point is the optimum without the level, 484.55, which is itself a point of the unrestricted model. Neither point is shown to be a global minimum, so the statistics are differences at the points found, not bounds. In the two demographic specifications the unrestricted estimate of the log income coefficient lies on the boundary of the parameter space, so the χ^2 critical value is an approximation. Panel B reports objective values only: in each of these specifications the version without the log income level attains a lower objective than any found with it. Objectives are comparable only within a specification.

B.5 Moments, weights, and standard errors

Stack the moments as $m(\theta) = (g(\theta)', d(\theta)')'$ with weight $W = \text{blkdiag}(nS_A^{-1}, V_M^{-1})$, so that the objective (15) is $m'Wm$. The aggregate block uses the IV scores $\psi_{jmt} = Z_{jmt}\xi_{jmt}$, centered at their mean. Its covariance is clustered two ways, $S_A = S_{\text{prod}} + S_{\text{DMA}} - S_{\text{prod} \times \text{DMA}}$, where each term is $\sum_c U_c U_c' / n$ for the sums U_c of the centered scores within the clusters c of that type, with no finite-sample correction. Negative eigenvalues of S_A , after scaling by the norms of the instruments, are set to zero, and its inverse drops eigenvalues below 10^{-10} times the largest. The household block uses V_M , the full covariance matrix of the vector of target estimates across 500 whole-household bootstrap resamples (Appendix D), so the covariance between moments that summarize overlapping purchases enters the weight. The two blocks come from separate samples and are independent.

Estimation is two-step. S_A is evaluated at a first-step estimate and held fixed while Q is minimized, and the reported estimates minimize Q under these weights.

The covariance of the estimates is $V = (G'WG)^{-1}$ with S_A re-evaluated at the estimate. G stacks the derivatives of the macro moments with respect to the nonlinear parameters, which include the response of the concentrated linear parameters, the derivatives of the macro moments with respect to the linear parameters, and the derivatives of the micro moments. This is the asymptotic covariance of efficient two-step GMM, the form PyBLP (Conlon and Gortmaker, 2020) reports under an updated weighting matrix; the estimates are computed with weights frozen at the first-step point, and the general sandwich $(G'W_fG)^{-1}G'W_f\Omega W_fG(G'W_fG)^{-1}$ with those weights W_f coincides with it asymptotically when the model is correctly specified and the first-step point is consistent. Because W already contains n , no further scaling is applied. The price coefficient α in the tables is concentrated under the fitting weights and its standard error comes from the same V . For a standard deviation $\sigma = \exp(\ell)$, the delta method gives $\text{Var}(\sigma) = \sigma^2 \text{Var}(\ell)$. Simulation error from the fixed draws is not added.

B.6 Optimization

I minimize Q with the covariance matrix adaptation evolution strategy (CMA-ES) from several starting points, including the estimates of neighboring specifications. I refine the best point with a trust-region Levenberg–Marquardt method on the stacked moments, using a finite-difference Jacobian. Integration uses 100 fixed simulation draws per DMA, shared by its retailers and fixed across months, so the objective is a deterministic function of the parameters. The Jacobian G in the standard errors is taken with respect to the nonlinear and linear parameters, so that the standard error of α reflects its concentration.

C Counterfactual Details

C.1 Scored and fixed products

Let $\mathcal{I}_{mt}$ be the products whose quantities are predicted and scored in market m and month t , and let $\mathcal{F}_{mt}$ contain the other products observed in the category, whose observed gallons are held fixed. A product is scored when the model supplies its utility and product group, which requires that it was sold somewhere by March 2015 (the set of model products changes over the estimation window; Appendix Table A.3), and when its 2015 price passes two checks: the price is at most \$50 per gallon, and for a product with local sales in January–March 2015, its first-quarter gallon-weighted price is also at most \$50 and the log difference between the two prices is at most 0.5. A product that fails a check is held fixed at its observed gallons rather than dropped. Two cases arise: products in the estimated set whose 2015 price fails the checks or that lack the price history the forecast needs, and products sold during the withdrawal that are not among the estimated products, new listings among them. Both are held at their observed gallons; a brand outside the estimation products with no product group before the withdrawal takes the group of its own first-quarter 2015 sales (Appendix A). Because the product definition is fixed before the withdrawal, no model receives a product coefficient estimated from withdrawal-period sales. Table C.1 reports the rival gallons in each class and the gallons of the products held fixed, by reason; 97.8 percent of rival gallons in Blue Bell’s markets are scored.

C.2 The forecast mean utility

The forecasts carry each product’s mean utility forward from the estimation window. For each specification, let $y_{jmt} = \delta_{jmt} - \alpha p_{jmt} - \tau_{\text{month}(t)}$ be its mean utility net of the price and month-of-year terms through March 2015. I project y_{jmt} on product, DMA–retailer, and year effects, together with the retailer-by-private-label effects in the specifications that have them, and call the remainder $\hat{\xi}_{jmt}$. The forecast mean utility in May–August 2015 is

$$\hat{\delta}_{jmt} = \hat{a}_j + \hat{c}_m + \hat{\phi}_{2015} + \hat{\kappa}_{jm} + \alpha p_{jmt} + \tau_{\text{month}(t)} + \bar{\xi}_{jm}, \quad (\text{C.1})$$

where $\hat{\phi}_{2015}$ is the year effect estimated from January–March 2015, $\hat{\kappa}_{jm}$ is the retailer-by-private-label effect (zero for branded products and for specifications without these effects), p_{jmt} is the observed 2015 price, and $\bar{\xi}_{jm}$ is the product’s local residual demand. The local residual follows an ordered rule: the mean of $\hat{\xi}_{jmt}$ over January–March 2015 in market m ; if the product has no sales there in those months, the mean over its earlier months in m ;

Table C.1: Rival gallons by history class and by reason for fixing

	Blue Bell markets		All markets	
	Gallons	Percent	Gallons	Percent
<i>By source of the local residual</i>				
Local sales in January–March 2015	29,661,649	97.3	48,729,028	97.8
Earlier local sales only	224,458	0.7	353,590	0.7
New to the market, sold elsewhere (zero residual)	545,151	1.8	668,477	1.3
Never sold anywhere by March 2015	48,497	0.2	82,674	0.2
<i>By treatment</i>				
Scored	29,817,804	97.8	48,617,151	97.6
Fixed by the price checks	613,453	2.0	1,133,944	2.3
with local January–March 2015 sales	588,384		1,098,342	
with earlier local sales only	2,718		3,555	
new to the market	22,351		32,046	
Fixed, utility or product group not supported	48,497	0.2	82,674	0.2
Total rival gallons	30,479,755	100.0	49,833,769	100.0

Observed May–August 2015 gallons of rival products, excluding Blue Bell, under the main RC’s forecasts with observed prices. Blue Bell markets are the DMA–retailer markets with high, medium, or low exposure; all markets add those without Blue Bell scanner sales. The history classes and the fixed volume depend only on the data and the scoring rule, not on the model, and the other specifications use the same rule. The products whose utility or group is not supported are exactly those never sold anywhere by March 2015. Appendix C defines the classes and the price checks.

and zero if it has no history in m but was sold elsewhere. Every model receives the same information under the same rule. Blue Bell is given its last observed price in the market in January–March 2015, and its mean utility follows equation (C.1); the evaluation design uses this Blue Bell utility and this treatment of the fixed products.

The forecast reserves observed volume for the fixed products. Let Q_{mt}^F be their observed gallons and M_{mt} the market size. If the demand system includes a modeled subset of these products, such as an estimated product whose 2015 price fails the checks, with aggregate predicted share P_{mt}^F , the scored-product quantities are

$$\hat{q}_{jmt} = (M_{mt} - Q_{mt}^F) \frac{s_{jmt}}{1 - P_{mt}^F}, \quad j \in \mathcal{I}_{mt}, \quad (\text{C.2})$$

and the outside share is adjusted by the same denominator, after aggregation over consumer types. The comparison thus conditions on the realized volume of the fixed products and evaluates the responding subset, the scored products and the outside option; it differs from a counterfactual for the whole market, in which the fixed products would also respond to Blue Bell’s absence. Let $s^I = \sum_{j \in \mathcal{I}} s_j$ and s_0 be the aggregate scored-inside and outside

probabilities from a menu that also contains the modeled part of $\mathcal{F}$. If $s^I + s_0 + P^F = 1$, then (C.2) implies

$$\sum_{j \in \mathcal{I}} \hat{q}_j + \hat{q}_0 = M - Q^F, \quad (\text{C.3})$$

which checks the reservation of fixed volume. It attributes no modeled appeal to the unmodeled products, and the outside quantity is measured in gallons of purchase opportunities rather than in consumers.

In the direct forecast, each forecast is computed with Blue Bell present and with it removed, and the model’s removal effect on a group d of rival products is

$$M_{dmt}^{direct} = \sum_{j \in d} (\hat{q}_{jmt}^{removed} - \hat{q}_{jmt}^{present}). \quad (\text{C.4})$$

This effect depends on two forecasts besides the substitution pattern: the rivals’ forecast 2015 shares, among which Blue Bell’s demand is divided, and the model’s forecast of Blue Bell’s own 2015 sales, which sets how much demand there is to divide. The rescaling of equation (17) removes the scaling by the second, and the inversion below removes the first. Table 10 in Section 6.5 reports each model’s misplaced share at the three steps.

C.3 Residual rules for the direct forecast

Table C.2 replaces the first-quarter rule of the direct forecast with four others, each applied to every product including Blue Bell: the mean of all local residuals from 2011 through March 2015, the local mean in the same month of 2014, the mean over October–December 2014, and the mean over December 2014–February 2015, which leaves out March 2015, the month of the first recall announcement. The misplaced shares move by up to 0.07, the five models keep the same order under all five rules, and the PL-pooled nested logit has the smallest under each, between 0.23 and 0.27. Among these five models, the RC model forecasts the smallest Blue Bell volume under every rule, 0.46 to 0.62 of Blue Bell’s 2014 gallons, and the December–February window, which drops March and adds December, lowers that volume rather than raising it, so the residual window does not explain why it is small. Each row uses the market–months in which every model’s forecast is computed under that rule, so the order within a row compares the models on common market–months, while differences across rows also reflect small changes in the sample. The check is limited to the residual, because the parameters are estimated on data through March 2015 and Blue Bell keeps its last first-quarter price under every rule.

The level the direct forecast predicts for each rival is the other input that the evaluation design removes. In markets without Blue Bell, where nothing was removed, each model’s

Table C.2: The direct forecast under five residual rules

Residual carried to May–August 2015	Logit	NL, PL separate	NL, PL pooled	RC	NL, PL separate, novelties in one nest
<i>A. Misplaced share of Blue Bell's volume</i>					
January–March 2015 (main)	0.464	0.381	0.253	0.540	0.429
All months, 2011–March 2015	0.503	0.368	0.259	0.583	0.410
Same month of 2014	0.397	0.351	0.233	0.470	0.368
October–December 2014	0.435	0.372	0.267	0.576	0.424
December 2014–February 2015	0.455	0.387	0.243	0.560	0.437
<i>B. Blue Bell volume, share of 2014 gallons</i>					
January–March 2015 (main)	0.836	0.776	0.670	0.623	0.663
All months, 2011–March 2015	0.736	0.746	0.647	0.461	0.662
Same month of 2014	0.865	0.786	0.652	0.617	0.699
October–December 2014	0.843	0.784	0.623	0.458	0.651
December 2014–February 2015	0.838	0.784	0.673	0.534	0.667

First step of the design (Table 10): direct forecast with observed 2015 rival prices and Blue Bell at its last first-quarter 2015 price. RCNL and the group RC are not computed under the alternative rules. Each rule sets the local residual of every product, Blue Bell included; a product without a local residual in the window takes the mean of all its local residuals, and zero if it has no local history. Market–months in which any model’s forecast fails under a rule are dropped for all models, leaving 1,200 to 1,224 market–months with 2014 Blue Bell sales. Blue Bell volume is the forecast’s Blue Bell gallons over the same months’ 2014 gallons. Because the first row uses this table’s common market–months, it can differ in the third decimal from Table 10, whose first step uses the 1,234 market–months in which every model’s direct forecast solves.

error in log rival sales falls with the rival’s price change from the first quarter to the summer, and more steeply the more price-sensitive the model: the gallon-weighted slope within market–months is -0.5 for logit and -2.2 to -3.7 for the RC model and the nested logits. For the rivals sold in the same market and month of 2014, 90 percent of rival gallons, replacing the first-quarter residual with each rival’s own 2014 utility moved by its price change since then lowers the gallon-weighted mean absolute log error by 20 to 28 percent and the slope by 39 to 54 percent for every model. On Blue Bell’s market–months the same replacement raises the nested logits’ forecast of Blue Bell’s own volume by 15 to 33 percent and their raw rival gains with it, by up to 30 percent for the PL-pooled nested logit, while the share of the removal effect that stays among the rivals moves by about 5 percent or less and the misplaced share by at most 0.04. The level errors of the direct forecast are thus sensitive to the timing of the residual and remain associated with price changes, which is why the evaluation design takes rival sales from the data and rescales each removal effect to Blue Bell’s 2014 volume; the slope is a diagnostic of the forecast errors, not an estimate of elasticities, because each forecast contains the model’s own price term and realized price changes can move with unforecast demand.

C.4 Returning Blue Bell to the observed 2015 market

For each model and market-month, the evaluation design starts from the direct forecast's inputs: the scored and fixed products, their observed 2015 prices, the reserved volume of the fixed products, and the market size. It replaces the forecast mean utilities of the scored rival products by the values $\tilde{\delta}_{mt}$ that solve

$$q_{jmt}(\tilde{\delta}_{mt}; \mathcal{J}_{mt} \setminus \mathcal{B}) = q_{jmt}^{obs}, \quad j \in \mathcal{I}_{mt}, \quad (\text{C.5})$$

the inversion that recovers mean utilities from shares (Berry et al., 1995), computed separately for each market-month and each model with its own parameters, so that with Blue Bell absent the model reproduces every scored product's observed gallons. Each market-month's calculation checks that the world without Blue Bell reproduces the observed gallons: across the scored market-months, the largest relative gap between the model's and the observed rival gallons is 10^{-12} for logit, the RC model, and the nested logits, and every scored market-month passes the same check for RCNL and the group RC. Blue Bell's products are then returned with the mean utility of equation (C.1) at their last first-quarter 2015 price, and the model's gallons with Blue Bell present follow from the same share function. The removal effect of equation (16) is the difference between the two worlds by product, grouped into the six destinations. The rescaling of equation (17) divides by the model's own Blue Bell gallons in the market-month; a market-month in which a model predicts no Blue Bell volume keeps a zero effect, which applies to 441 of the 155,101 scored rival rows, the same rows in every model. The prediction error $\hat{Q}^{removed} - Q^{obs}$, the model's removal effect, and the observed year-over-year change have different reference states; the scores use only the removal effect, compared with the observed difference in differences.

C.5 Scores

Writing $e_d = m_d - o_d$ for the destination errors, $P = \sum_{d:e_d>0} e_d$ and $N = -\sum_{d:e_d<0} e_d$, the retention gap is $T = P - N$, the part of the destination errors that cancels in the total is $A = \min(P, N)$, and the misplaced share is $L = |T| + A = \max(P, N)$. Because A depends on the level of the removal effects, allocation at a fixed total is measured separately by the allocation error of equation (21). When both normalized gain vectors have nonnegative entries, as in the point estimates, the allocation error lies between zero and one and reads as a share of the rival gain. Adding the remainder as a seventh destination, with model and observed values $1 - \sum_d m_d$ and $1 - \sum_d o_d$, gives $L = \frac{1}{2} \sum_{d=0}^6 |m_d - o_d|$. In the model the remainder is the outside option; on the observed side it also contains new listings (about

0.13 per Blue Bell gallon, Table 3), other fixed products, and any error in the trend. The misplaced share reads as the share of Blue Bell’s volume placed in the wrong destination when every destination effect is nonnegative and their sum is below one, as in the point estimates; in bootstrap draws where an observed effect is negative or the observed total exceeds one, the statistics keep their definitions and nothing is truncated.

Table C.3 works through the scores for one market–month with invented round numbers. Blue Bell sold 100 gallons in the same month of 2014, and the observed removal effects on three rivals are 40, 25, and 10 gallons, so 75 gallons stayed with the scored rivals and 25 are the remainder. Model 1 predicts effects of 50, 15, and 5 gallons. Its retention, 0.70, is close to the observed 0.75, so its retention gap is -0.05 , but it places too much on rival A; its destination differences are 0.10, -0.10 , and -0.05 , its allocation error is 0.181, and its misplaced share is $\frac{1}{2}(0.25 + 0.05) = 0.15$. Model 2 predicts 30, 20, and 5 gallons. It keeps the rivals in about the observed proportions, with an allocation error of 0.042, but retains only 0.55, so its retention gap is -0.20 and its misplaced share 0.20. Adding the remainder as a fourth destination gives the same misplaced share as half the sum of absolute differences over all four destinations; because every effect here is nonnegative, it reads as the share of Blue Bell’s volume that each model places in the wrong destination. The example shows why I report the three statistics together: the misplaced share ranks Model 1 first, while the two margins show that the models fail in different ways.

Table C.3: Hypothetical illustration of the scores (invented numbers, not results)

Destination	Removal effect, gallons			Difference per Blue Bell gallon	
	Observed	Model 1	Model 2	Model 1	Model 2
Rival A	40	50	30	0.10	-0.10
Rival B	25	15	20	-0.10	-0.05
Rival C	10	5	5	-0.05	-0.05
All rivals (retention $\times 100$)	75	70	55		
Remainder	25	30	45	0.05	0.20
Retention gap (model minus observed)				-0.05	-0.20
Allocation error (share of the rival gain)				0.181	0.042
Misplaced share of Blue Bell’s volume				0.15	0.20

One market–month in which Blue Bell sold 100 gallons in the same month of 2014. All numbers are invented for exposition. Model 1 retains nearly the observed volume but places it on the wrong rivals; Model 2 places it in about the right proportions but retains too little.

All scores are pooled. I sum the effects over the scored market–months before dividing by their total 2014 Blue Bell gallons, which weights each market–month by Blue Bell’s volume in it and lets errors cancel across market–months, so the score answers where Blue Bell’s demand went in total. A Blue Bell-volume-weighted average of losses computed market–

month by market-month would also weight by volume but would not let errors cancel; it answers how well each market-month is predicted, a different question that I do not report.

Sampling uncertainty comes from resampling DMAs. Each draw resamples, with replacement, the DMAs of Blue Bell’s markets and of the comparison markets, with their retailers and months, and recomputes the comparison ratios r_d , the observed effects, and every model’s effects at the estimated parameters. Because the draws are common to all models, differences between models are paired, and Appendix Table C.4 reports the share of draws in which one model has the smaller score. An unmodeled shock common to all of Blue Bell’s markets in 2015 would move every model’s error together.

Table C.4: Paired comparisons: share of draws in which the row model has the smaller score

		NL, Logit	NL, PL separate	NL, PL pooled	RC	RCNL	Group RC
<i>A. Misplaced share</i>							
Logit	.		0.03	0.05	0.01	0.04	0.01
NL, PL separate	0.97		.	0.07	0.89	0.05	0.97
NL, PL pooled	0.95		0.93	.	0.94	0.54	0.95
RC	0.99		0.11	0.06	.	0.05	0.99
RCNL	0.96		0.95	0.46	0.95	.	0.96
Group RC	0.99		0.03	0.05	0.01	0.04	.
<i>B. Absolute retention gap</i>							
Logit	.		0.02	0.03	0.01	0.03	0.01
NL, PL separate	0.98		.	0.13	0.97	0.11	0.98
NL, PL pooled	0.97		0.87	.	0.96	0.63	0.96
RC	0.99		0.03	0.04	.	0.04	0.99
RCNL	0.97		0.89	0.37	0.96	.	0.97
Group RC	0.99		0.02	0.04	0.01	0.03	.
<i>C. Allocation error</i>							
Logit	.		0.97	0.53	0.77	0.25	0.79
NL, PL separate	0.03		.	0.05	0.03	0.03	0.03
NL, PL pooled	0.47		0.95	.	0.60	0.33	0.62
RC	0.23		0.97	0.40	.	0.14	0.81
RCNL	0.75		0.97	0.67	0.86	.	0.87
Group RC	0.21		0.97	0.38	0.19	0.13	.

Share of the paired DMA-bootstrap draws in which the row model has the smaller score than the column model, evaluation design of Section 6.1: the misplaced share in panel A, the absolute value of the retention gap in panel B, and the allocation error in panel C. A dot marks the diagonal.

D Household Benchmark Construction

D.1 The observed buyer-cohort benchmark

Let $\mathcal{H}$ contain the households that bought Blue Bell in the selection year and are on the panel roster in the selection year, the base year, and the event year. For a rival brand b , let $G_{b,14}$ and $G_{b,15}$ be gallons purchased by those households in the base and event windows, and let $G_{BB,14}$ be their Blue Bell gallons in the base window. With reference trend g_b , the net rival increment is

$$\Delta G_b = G_{b,15} - g_b G_{b,14}. \quad (\text{D.1})$$

Net retention and the destination composition are

$$R = \frac{\sum_b \Delta G_b}{G_{BB,14}}, \quad c_b = \frac{\Delta G_b}{\sum_k \Delta G_k}. \quad (\text{D.2})$$

The main benchmark uses modeled products and a common trend g estimated from households in the same DMAs that bought no Blue Bell in the selection or base year; brand-specific trends g_b and the all-product rival set are sensitivity choices.

The benchmark uses 2013 as the selection year and May–August 2014 and 2015 as the base and event windows, with a cohort of 5,429 households, 5,147 of whom bought ice cream in May–August 2014, and 22,280 reference households. The estimated common trend is 0.964, and within modeled products R is 0.555 with a standard error of 0.043. A cohort household that bought no Blue Bell in May–August 2014 adds nothing to $G_{BB,14}$ but contributes its trend-adjusted rival change to the numerator. The destination shares in Table 5 use the same cohort, trend, and products.

The cohort is selected on the year before the base window because selecting households on a Blue Bell purchase in the base window would favor households whose purchases in that window were high by chance, which mixes the removal with regression to the mean. The same construction applied to 2012 buyers between May–August 2013 and May–August 2014 gives a net retention of -0.017 , so it produces no retention when nothing is removed. The models are not scored against this benchmark because its cohort is selected on realized purchases, which a model counterpart would have to replicate on simulated draws before comparing quantities.

Table D.1 compares this benchmark with the one obtained when the cohort is selected on Blue Bell purchases in May–August 2014, the window that also measures the volume removed. That cohort has 4,442 households, its net retention is 0.475 with a standard error of 0.020, and its placebo is 0.006. The destination shares of the two cohorts are within 0.02

of each other for every brand.

Table D.1: The household benchmark under two cohort definitions

	Bought Blue Bell in 2013	Bought Blue Bell in May–August 2014
Base-window ice cream buyers	5,147	4,442
Comparison households	22,280	26,109
Common trend $\hat{g}$	0.964	0.960
Net retention, estimation products	0.555 (0.043)	0.475 (0.020)
Net retention, all ice cream	0.595 (0.046)	0.503 (0.022)
Placebo one year before	−0.017	0.006
<i>Share of the net rival increment, estimation products</i>		
Private label	0.416 (0.036)	0.414 (0.028)
Blue Bunny	0.161 (0.019)	0.170 (0.017)
Breyers	0.094 (0.016)	0.092 (0.014)
Dreyer’s	0.078 (0.015)	0.067 (0.014)
Ben & Jerry’s	−0.002 (0.006)	0.005 (0.004)
Häagen-Dazs	0.017 (0.008)	0.012 (0.005)
Other brands	0.237 (0.028)	0.240 (0.023)

Both columns measure the volume removed as the cohort’s May–August 2014 Blue Bell gallons and the response as the trend-adjusted change in rival gallons from May–August 2014 to May–August 2015. The first column is the cohort of Table 5. The second selects households that bought Blue Bell in May–August 2014 and stayed on the panel roster through 2015; its comparison households are households in the same DMAs with no Blue Bell purchase in 2013 or 2014, and its placebo uses 2013 buyers between 2013 and 2014. Standard errors from 50 whole-household bootstrap resamples.

The household bootstrap resamples complete households and recomputes the trend and destination statistics with the same rule as the point estimate, so the standard errors reflect uncertainty in both the ratio and the trend adjustment. When a net increment is negative, a destination’s c_b can be negative; it is then a signed contribution to a net change rather than a probability of individual switching.

D.2 Robustness and detail of the benchmark

Table D.2 reports the benchmark under alternative trends. Net retention within the estimation products is 0.490 without a trend adjustment, 0.580 with a trend of 0.95, 0.540 with brand-specific trends, and 0.606 with the trend of households in the DMAs without Blue Bell sales, against 0.555 under the common trend. The destination shares are stable across these trends: Blue Bunny stays between 0.15 and 0.17, private label between 0.41 and 0.42, and other brands between 0.22 and 0.24.

The same-year non-purchase comparison of Section 3 uses the control households reweighted to the cohort’s 2014 category volume. The unweighted control is not informative

Table D.2: The household benchmark under alternative trends

	Common $\hat{g} = 0.964$	$g = 1$	$g = 0.95$	Brand trends	Zero-DMA trend
Net retention, estimation products	0.555	0.490	0.580	0.540	0.606
Net retention, all ice cream	0.595	0.517	0.625	0.583	0.655
Placebo: 2012 buyers, 2013 to 2014	-0.017	-0.085	0.001	0.001	.
<i>Share of the net rival increment, estimation products</i>					
Private label	0.416	0.410	0.418	0.417	0.420
Blue Bunny	0.161	0.171	0.157	0.174	0.154
Breyers	0.094	0.098	0.092	0.075	0.091
Dreyer's	0.078	0.082	0.076	0.091	0.075
Ben & Jerry's	-0.002	-0.006	-0.001	0.002	0.000
Häagen-Dazs	0.017	0.017	0.017	0.019	0.017
Other brands	0.237	0.228	0.240	0.223	0.243

Same cohort, windows, and products as Table 5. The common trend comes from households in the same DMAs that bought no Blue Bell in 2013 or 2014; $g = 1$ applies no trend; brand trends use each brand's own trend among those households; the zero-DMA trend comes from households in the DMAs without Blue Bell sales. A dot marks a statistic not computed for that trend.

for this comparison: its rate, 0.115, lies above the cohort's 0.095, because households that bought no Blue Bell are lighter ice cream buyers and reweighting them to the cohort's volume lowers the rate to 0.101.

Table D.3 decomposes the volume that rival purchases do not replace by the households' 2015 purchase status.

Table D.3: Unretained Blue Bell volume by 2015 purchase status

	Recall cohort 2014 to 2015	Placebo cohort 2013 to 2014
Unretained volume, one minus net retention	0.445 (0.043)	1.017 (0.035)
Households with no ice cream purchase	0.124 (0.010)	0.093 (0.008)
Their Blue Bell volume	0.066 (0.006)	0.040 (0.005)
Their trend-adjusted rival volume	0.058 (0.006)	0.053 (0.006)
Continuing households	0.322 (0.041)	0.924 (0.034)
Normal-exit benchmark	0.115 (0.005)	0.102 (0.005)
Excess non-purchase	0.009 (0.010)	-0.009 (0.008)
Households	5,429	5,257
Buying ice cream in the base window	5,147	4,960

Volumes per gallon of the cohort's Blue Bell purchases in the base window, estimation products, common trend; standard errors from the 50 whole-household bootstrap resamples of Table 5. Unretained volume splits into the households with no ice cream purchase in the event window and the continuing households. The normal-exit benchmark is the volume that non-purchase at the rate of the comparison households in the same DMAs, reweighted in the five base-volume bins of the exit comparison, would remove, and excess non-purchase is the difference between the no-purchase volume and the benchmark. Non-purchase rates are measured among the households that bought ice cream in the base window.

Table D.4 splits the 5,147 cohort households that bought ice cream in May–August 2014 by their 2013 Blue Bell purchases, at 1.5 and 2.5 gallons, into 3,364 light, 541 medium, and 1,242 heavy buyers. Net retention is similar across the tiers, 0.578, 0.609, and 0.537, but the standard errors of the light and medium tiers are more than three times the heavy tier’s, because light buyers bought little Blue Bell and the medium tier is small. Blue Bunny’s share of the net increment relative to its baseline share is 2.7 among light and medium buyers, against 1.4 among heavy buyers. The heavy tier is the most precisely measured. Blue Bunny takes 0.128 of its increment, private label 0.448, and other brands 0.245.

Table D.4: Net retention, allocation, and non-purchase by buyer tier

	Light	Medium	Heavy
Households	3,364	541	1,242
Net retention	0.578 (0.122)	0.609 (0.106)	0.537 (0.035)
Blue Bunny share of net rival increment	0.219 (0.062)	0.210 (0.052)	0.128 (0.017)
Blue Bunny share of 2014 rival basket	0.082	0.078	0.094
Ratio	2.7	2.7	1.4
Private-label share of net rival increment	0.365 (0.090)	0.356 (0.092)	0.448 (0.030)
Other brands’ share of net rival increment	0.213 (0.088)	0.242 (0.060)	0.245 (0.025)
Share buying no ice cream in May–August 2015	0.105 (0.005)	0.074 (0.010)	0.076 (0.008)
Ratio to same-year reweighted control	0.92	0.75	1.11

Tiers split the 5,147 households of Table 5 that bought ice cream in May–August 2014 by their 2013 Blue Bell purchases, at 1.5 and 2.5 gallons. Same construction and common trend as Table 5; standard errors from 50 whole-household bootstrap resamples. Non-purchase is measured among households that bought ice cream in May–August 2014; the same-year control is households in the same DMAs that bought no Blue Bell in 2013 or 2014, reweighted to the cohort’s 2014 category volume.

Non-purchase is highest among light buyers, 0.105, against 0.074 among medium and 0.076 among heavy buyers. Relative to the same-year control, the ratios are 0.92, 0.75, and 1.11, with no ordering by tier.

Table D.5 splits the cohort by when Blue Bell returned to its DMA and extends the event window. In the DMAs where Blue Bell was back by November 2015, net retention was 0.563 (standard error 0.048), against 0.554 (0.051) in the late-return DMAs, and Blue Bunny took 0.116 and 0.187 of the increment. Extending the window for each DMA from May to the month before its own return gives 5,337 cohort households that bought ice cream in the longer base window, against 5,147 in May–August; over this window net retention is 0.581 (0.038), with Blue Bunny at 0.137, private label at 0.363, other brands at 0.310, and a non-purchase ratio to the same-year control of 0.99. The longer window changes the households, the months, and the markets together, so its difference from the May–August benchmark is descriptive.

Blue Bell regained a substantial share of its former buyers’ purchases while its product

Table D.5: Net retention and allocation by return timing and window length

	May–August window		Window length	
	Early-return DMAs	Late-return DMAs	May–August, all DMAs	May to each DMA’s return
Base-window ice cream buyers	1,493	3,647	5,147	5,337
Net retention	0.563 (0.048)	0.554 (0.051)	0.555 (0.043)	0.581 (0.038)
Blue Bunny share	0.116	0.187	0.161	0.137
Private-label share	0.390	0.428	0.416	0.363
Other brands’ share	0.266	0.221	0.237	0.310
Non-purchase ratio to same-year control	1.14	0.85	0.94	0.99

Same cohort rule, construction, and common trend as Table 5; shares are shares of the net rival increment, estimation products; standard errors from 50 whole-household bootstrap resamples. Early-return DMAs are those where Blue Bell was back by November 2015. The last column uses, in both years, the months from May to the month before Blue Bell’s return to the household’s DMA, through December. The non-purchase control is reweighted as in Table D.4.

coverage remained limited. The return series follows households that bought Blue Bell during 2014, stayed on the panel roster from 2014 through 2016, and lived in DMAs where Blue Bell returned by June 2016, with event time measured from each DMA’s return. Among them, Blue Bell’s share of category volume was 0.25 in the month the brand returned to their DMA, 62 percent of its 0.40 in the same calendar month of 2014, and 0.32, 82 percent, by the eleventh month, while the scanner data show only 20 to 35 percent of its 2014 product–markets back on the shelf. The recovery after the return differs by tier. Light buyers returned to 80 to 100 percent of their 2014 Blue Bell share within a few months, whereas heavy buyers stayed at 65 to 75 percent through the twelfth month. Among the households on the roster in all four years, Blue Bell’s share of category volume was 0.32 in 2013 and 0.35 in 2014, fell to 0.14 in 2015, and reached 0.19 in 2016; in the second half of 2016 it remained a little below 60 percent of its 2013 level. Between May–August 2014 and May–August 2015 these households’ private-label share rose from 0.25 to 0.38 and their Blue Bunny share from 0.06 to 0.11.

With Blue Bell at 65 to 80 percent of its 2014 share after the return, the rivals together sit about 15 to 20 percent above their 2014 shares (Table D.6). Private label, at 1.18 times its 2014 share, and other brands, at 1.12 to 1.15, are near that level. Blue Bunny, at 1.56 in the first six months after the return and 1.23 in the next six, and Breyers, at 1.4, are above it, while Dreyer’s is at about its 2014 level. A year after the return, the cohort’s purchases still tilted toward the mainstream branded tubs.

Table D.6: Blue Bell’s buyers after the return

<i>Blue Bell’s share of the cohort’s category volume, months since each DMA’s return</i>	
Return month (same calendar month of 2014: 0.40)	0.25
Eleventh month	0.32
Share of Blue Bell’s 2014 product–markets back on the shelf, same months	0.20 to 0.35
Light buyers, share relative to 2014, first months	0.80 to 1.00
Heavy buyers, share relative to 2014, through month 12	0.65 to 0.75
<i>Rivals’ shares relative to their 2014 shares after the return</i>	
Private label	1.18
Other brands, months 0 to 5 / 6 to 11	1.12 / 1.15
Blue Bunny, months 0 to 5 / 6 to 11	1.56 / 1.23
Breyers	1.4
Dreyer’s	about 1.0
<i>Blue Bell’s annual share of category volume, households on the roster all four years</i>	
2013 / 2014 / 2015 / 2016	0.32 / 0.35 / 0.14 / 0.19

The return series follows households that bought Blue Bell during 2014, were on the roster in 2014–2016, and belonged to 2014 Blue Bell DMAs with a detected return by June 2016. Relative shares are averaged over event months 0–5 and 6–11, retaining months observed in at least 30 DMAs. Event time is measured from each DMA’s own return month, and the annual shares in the last panel use households on the roster in all four years. Availability is the fraction of Blue Bell’s 2014 product–markets with scanner sales in the same month. Rivals’ ratios compare their share of the cohort’s category volume after the return with the same share in 2014.

D.3 Purchase occasions and annual events

The model maps the micro moments to monthly choices. A simulated consumer i keeps its type within a DMA-year k and makes one choice in each month $m = 1, \dots, 12$, and products of different groups are exclusive within a month. Let $p_{im}(B)$ be the probability that consumer i buys a product in the set B in month m , where B is the inside goods, the products with a given characteristic, or the products of a given group. With independent choice shocks across months, consumer i ’s expected number of purchases in B during the year and its probability of buying in B at least once are

$$L_i(B) = \sum_{m=1}^{12} p_{im}(B), \quad A_i(B) = 1 - \prod_{m=1}^{12} (1 - p_{im}(B)). \quad (\text{D.3})$$

Quantity moments are ratios of sums rather than averages of individual ratios. Within a DMA-year, the mean quantity among buyers of B is $\sum_i L_i(B) / \sum_i A_i(B)$ over the simulated

consumers, and the moment combines markets with weights w_k equal to their market size:

$$m(B) = \frac{\sum_k w_k E_i[L_i(B)]}{\sum_k w_k E_i[A_i(B)]}, \quad (\text{D.4})$$

in gallons, since a model purchase is one gallon: the market size is in gallons, so each consumer's monthly choice is a one-gallon purchase. The observed quantity targets are in gallons, and the price moments in dollars per gallon. Equivalently, each market's quantity among buyers is weighted by its market size times its buying probability. The mean quantity of buying households uses the inside goods as B , the conditional quantities use the products with each characteristic, and the income-group moments evaluate (D.4) within the three income groups that also define the income shifts. The own-group moments apply (D.4) to the products of each group g .

The cross-group rows measure how much households that buy one group buy of another. No specification I report targets them. For an ordered pair of groups $g \neq h$, the moment is the expected number of group- h purchases in the year among consumers who bought group g at least once:

$$m_{gh} = \frac{\sum_k w_k E_i[N_{ih}\mathbf{1}(N_{ig} \geq 1)]}{\sum_k w_k E_i[A_i(g)]}, \quad E[N_{ih}\mathbf{1}(N_{ig} \geq 1)] = \sum_m p_{im}(h) \left[1 - \prod_{s \neq m} (1 - p_{is}(g))\right], \quad (\text{D.5})$$

where N_{ig} is consumer i 's number of group- g purchases in the year. The second expression follows because a month with a group- h purchase has no group- g purchase, so g must be bought in one of the other eleven months. The four groups give twelve ordered pairs.

The observed targets count purchases in gallons over a year, a household enters the conditioning set of a moment when it buys at least one gallon of the group in the year, and it can buy several packages or brands on one trip, while the model allows one one-gallon choice per month and conditions on at least one such purchase. The two buying events therefore differ, and the difference matters most for moments conditioned on buying and for event probabilities near one. The target file's two portfolio moments, the probability of buying any novelty among households that buy any non-novelty product and the reverse, are dropped for this reason: at every parameter vector evaluated the model gives 0.97 to 0.98 against a target of 0.895.

Table D.7 records the observed statistic and its model counterpart for every block.

Table D.7: Micro moment definitions

Block	Observed statistic	Model counterpart
Base (12)	Mean quantity of buying households; mean quantity given purchase of each of the seven characteristics; middle and high minus low income group in quantity and in average price paid	Equation (D.4) for the inside goods, for each characteristic, and by income group; price paid from the same probabilities
Income-group contrasts (8)	Middle and high minus low income group in premium, novelty, and private-label volume shares and in price paid	Ratios of expected volumes and expenditures by income group, from the same monthly probabilities
Own-group (4)	Annual purchases within each group among households that buy the group, in gallons	Equation (D.4) for the products of each group
Cross-group (12)	Annual purchases in group h among households that buy group g at least once, in gallons	Equation (D.5) for each ordered pair
Recall benchmark	Net rival increment of the 2013 buyer cohort	Same selection, baseline quantities, and trend rule

Targets are computed from 2011–2014 Homescan households; their covariance comes from 500 whole-household bootstrap resamples. Each simulated consumer makes one choice per month and keeps its type within a DMA-year; moments are ratios of sums over simulated consumers, combined across DMA-years with market-size weights.

D.4 Fit to the micro moments

Table D.8 reports observed micro moments and the available model counterparts, and Table D.9 summarizes fit by block and estimation window. The main and demographic RC specifications fit the eight base quantities within about 4 percent on average, and the group RC within 8 percent. The main and demographic RC specifications overpredict the quantity bought by buyers of cookies or mix-ins by about 11 percent, and the group RC underpredicts that of low-calorie buyers by 11 percent. The own-group moments are missed by about 6 to 8 percent on average. The cross-group moments are missed by about 30 percent on average, most of all premium-tub purchases among mainstream-tub buyers, which these specifications overpredict by 97 to 112 percent. The main and demographic RC specifications reproduce the sign of all twelve income contrasts and the group RC eleven.

Table D.8: Household micro moments: data and model

	Data	RC,			
		RC	demographic	RCNL	Group RC
<i>A. Base block (targeted by every RC specification)</i>					
All ice cream, among buying households	5.49	5.32	5.35	...	5.13
Vanilla, among buyers	1.90	1.85	1.86	...	1.71
Chocolate, among buyers	2.08	2.05	2.05	...	1.93
Bulk, among buyers	3.17	3.08	3.10	...	2.93
Low fat, among buyers	1.81	1.77	1.77	...	1.64
Low calorie, among buyers	1.87	1.81	1.83	...	1.66
Assorted, among buyers	1.76	1.73	1.74	...	1.66
Cookies or mix-ins, among buyers	1.07	1.18	1.18	...	1.16
Middle minus low income, quantity	-0.02	-0.06	-0.11	...	0.06
High minus low income, quantity	-0.28	-0.36	-0.37	...	-0.36
Middle minus low income, price paid	0.86	0.96	0.96	...	0.82
High minus low income, price paid	2.11	2.11	2.05	...	2.14
<i>B. Own-group purchases among the group's buyers (diagnostic as shown)</i>					
Mainstream tubs	2.77	2.89	2.90	...	2.75
Premium tubs	1.22	1.31	1.31	...	1.28
Novelties	1.67	1.56	1.58	...	1.48
Private label	2.90	2.68	2.70	...	2.50
<i>C. Income-group contrasts (targeted by the demographic RC)</i>					
Middle minus low income, premium share	0.010	0.013	0.012	...	0.014
High minus low income, premium share	0.021	0.029	0.028	...	0.033
Middle minus low income, novelty share	0.026	0.024	0.025	...	0.017
High minus low income, novelty share	0.049	0.044	0.044	...	0.035
Middle minus low income, private-label share	-0.065	-0.032	-0.034	...	-0.023
High minus low income, private-label share	-0.123	-0.042	-0.044	...	-0.032
Middle minus low income, price paid	0.84	1.00	0.99	...	0.89
High minus low income, price paid	1.99	2.13	2.07	...	2.22
<i>D. Cross-group purchases (diagnostic)</i>					
Premium tubs, among mainstream-tub buyers	0.18	0.39	0.39	...	0.36
Novelties, among mainstream-tub buyers	0.86	0.78	0.79	...	0.72
Private label, among mainstream-tub buyers	1.79	1.93	1.95	...	1.76
Mainstream tubs, among premium-tub buyers	2.19	2.71	2.73	...	2.47
Novelties, among premium-tub buyers	1.18	0.83	0.84	...	0.73
Private label, among premium-tub buyers	1.39	1.84	1.87	...	1.58
Mainstream tubs, among novelty buyers	2.32	2.11	2.14	...	1.87
Premium tubs, among novelty buyers	0.23	0.36	0.36	...	0.32
Private label, among novelty buyers	1.89	2.10	2.13	...	1.82
Mainstream tubs, among private-label buyers	1.96	1.84	1.86	...	1.71
Premium tubs, among private-label buyers	0.16	0.31	0.31	...	0.28
Novelties, among private-label buyers	0.82	0.75	0.76	...	0.69

Moments of 2011 to 2014 Homescan households at each specification's estimate. Quantities in gallons per household-year; volume shares as fractions of the income group's volume; price paid in dollars per gallon, 2019 dollars in panel C. A household buys a characteristic or a group when it buys at least one gallon of it in the year. The groups are those of the PL-separate grouping with the novelties in one group; the group RC targets the own-group moments on its PL-pooled groups instead, so panel B is diagnostic for every specification shown. The cross-group moments are targeted by no specification. ...: RCNL's moments at its estimate, not yet computed. Appendix Table D.7 gives the model counterpart of each block.

Table D.9: Fit to the household micro moments by block

		Base quantities (8)			Own-group (4)			Cross-group (12)		Income contrasts
Specification	Targets	Mean	Largest		Mean	Largest		Mean	Largest	Signs as in data
<i>Estimated on data through March 2015</i>										
RC (main)	B	3.7	Cookies	+10.7	6.4	Private label	−7.6	33.2	+110.0	12
RC without retailer ×	B	3.7	Cookies	+10.7	6.4	Private label	−7.7	33.1	+109.9	12
PL effects										
RC, demographic	B, C	3.1	Cookies	+11.0	6.2	Premium	+7.6	33.6	+111.5	12
RCNL	B	...			...			...		...
Group RC	B, O	8.4	Low calorie	−11.4	7.8	Private label	−13.7	29.3	+97.2	11
<i>Full 2011–2016 window</i>										
RC (main)	B	2.8	Cookies	+10.8	6.3	Premium	+8.6	.		12
RC without retailer ×	B	2.8	Cookies	+10.8	6.3	Premium	+8.6	.		12
PL effects										
RC, demographic	B, C	2.9	Cookies	+10.9	6.2	Premium	+8.6	.		12

“Mean” is the mean absolute percent gap between the model and the data over the block, and “Largest” names the moment with the largest absolute percent gap and gives its signed percent gap. Base quantities: the mean annual quantity of buying households and the quantity among buyers of each of the seven characteristics. Own-group and cross-group moments are defined on the PL-separate groups with the novelties in one group, against which every specification is compared, so they are diagnostic for the group RC as well. In every specification the largest cross-group gap is on premium-tub purchases among mainstream-tub buyers. The cross-group moments are evaluated for the specifications estimated on data through March 2015, and a dot marks a block not evaluated. The last column counts the twelve income contrasts (the four in the base block and the eight income-group contrasts) whose model value has the sign of the data; their levels are in Appendix Table D.8, and their percent gaps are not reported because several observed contrasts are close to zero. Targets: B the 12 base moments, C the eight income-group contrasts, O the four own-group moments on the PL-pooled groups. ...: RCNL's moments at its estimate, not yet computed. PL is private label.

E The Supply Side

The model-price route of Section 7 solves each model's prices with and without Blue Bell under multiproduct Bertrand pricing. This appendix sets out the pricing problem and the recovery of marginal costs, and it reports the diagnostics that decide where the route is defined.

E.1 Costs, ownership, and repricing

The price calculation uses multiproduct profit maximization with the recorded ownership of branded products and retailer ownership of private label. For an owner o and a free price p_j , the first-order condition is

$$q_j + \sum_{k \in \mathcal{J}_o} (p_k - c_k) \frac{\partial q_k}{\partial p_j} = 0. \quad (\text{E.1})$$

Baseline marginal costs are recovered from baseline prices and demand derivatives under this conduct assumption, so reproducing the baseline equilibrium is an accounting check rather than a test of conduct.

Let $J_{kj} = \partial q_k / \partial p_j$, and define the ownership-adjusted matrix B with row j , column k equal to $1\{o(j) = o(k)\}J_{kj}$. At an interior baseline equilibrium, the first-order conditions are

$$q + B(p - c) = 0, \quad c = p + B^{-1}q, \quad (\text{E.2})$$

when B is invertible. I report negative recovered costs, extreme margins, nonnegative own-price responses, and first-order-condition residuals as diagnostics and keep the recovered costs unclipped.

The ownership mapping assigns branded products to recorded owners and private label to its retailer in a DMA-retailer market. A price-solving unit may therefore differ from a brand. When only some prices are free, let $\mathcal{R}_o$ contain the owner's free-price products and $\mathcal{K}_o$ its fixed-price products. For $j \in \mathcal{R}_o$,

$$q_j + \sum_{k \in \mathcal{R}_o} (p_k - c_k) J_{kj} + \sum_{k \in \mathcal{K}_o} (p_k^{\text{fixed}} - c_k) J_{kj} = 0. \quad (\text{E.3})$$

The final term remains even though those prices are fixed, because the owner's objective includes the margins on its fixed-price products. A portfolio for which no cost can be recovered before the withdrawal keeps its observed prices.

I recover costs with the same demand construction used in the first-order conditions and

carry them into the forecast as the local January–March 2015 means, with the fallbacks of Appendix C.

E.2 Diagnostics of the recovered costs

Table E.1 reports, for each model, the own-price elasticity, the share of a price-induced loss that goes to the outside option, the implied Lerner index, and the share of private-label volume with a negative recovered marginal cost. The PL-separate nested logit recovers private-label margins above the price, as does its variant with novelties in one nest. Their volume-weighted private-label Lerner index is 1.83 and 1.85, and about 96 percent of private-label volume has a negative recovered cost, so I do not compute their model-price route. RCNL has a gallon-weighted median own-price elasticity of -4.2 and a volume-weighted Lerner index of 0.45, against 0.41 in the RC model and 0.38 in the PL-pooled nested logit, with 10.7 percent of its private-label volume at a negative recovered cost. The group RC’s Lerner index is 0.45, and 0.61 for private label, with 0.5 percent of private-label volume at a negative recovered cost. Logit recovers a negative cost for about a fifth of private-label volume, which the forecast keeps unclipped.

Table E.1: Demand and cost diagnostics behind the model prices

	Own-price elasticity	Share of loss to outside	Lerner index	Lerner index, private label	Private-label volume, negative cost
Logit	−1.87	0.49	0.62	0.85	0.209
NL, PL separate	−3.41	0.17	1.00	1.83	0.958
NL, PL pooled	−4.51	0.10	0.38	0.60	0.049
RC	−2.99	0.41	0.41	0.56	0.002
RCNL	−4.20	0.13	0.45	0.71	0.107
Group RC	−2.64	0.48	0.45	0.61	0.005
NL, PL separate, novelties in one nest	−4.66	0.13	0.97	1.85	0.959

The own-price elasticity is the gallon-weighted median of Table 7; the other columns are volume-weighted over rival products on the same products and markets. The Lerner index covers all rivals. The share of loss to outside is the share of the sales a product loses when its price rises that goes to the outside option. Marginal costs are recovered under multiproduct Bertrand pricing with recorded ownership and retailers pricing their own private label (equation (E.1)). The last column is the share of private-label volume whose recovered marginal cost is negative; costs are not clipped.

E.3 Welfare units

For nested logit conditional on type i , the inclusive value in (24) is

$$\mathcal{I}_i = \log \left(1 + \sum_g D_{ig}^{1-\rho} \right). \quad (\text{E.4})$$

The expected-utility constant cancels when comparing menus under the same error normalization. Fixed-price removal shrinks the menu and weakly lowers each type's inclusive value, and letting other prices change adds the effect of the price response. The money conversion divides by $-\alpha_i$, which requires every type's price coefficient to be negative.

If one modeled opportunity is q_0 gallons, the market contains M/q_0 opportunities, and multiplying per-opportunity surplus by this number gives aggregate dollars. Welfare is computed in the evaluation design's choice system, with the fixed products at their observed volume as in (C.2).

Table 11 reports the availability term in two units. Per gallon of the model's own Blue Bell volume, it divides the term in each market-month by the model's Blue Bell gallons. Per 2014 Blue Bell gallon, it weights each market-month's term per model Blue Bell gallon by V_{mt} , as equation (17) does for the removal effect. The price term, which does not scale with each model's Blue Bell volume, is summed over market-months and divided by Blue Bell's 2014 gallons.

In logit and nested logit the availability term is tied to retention. Let b be the share of purchase opportunities that Blue Bell takes when present, s_0^- the outside share without it, and R the share of Blue Bell's demand that stays with the rivals. Because the outside option is a singleton nest, the outside share with Blue Bell present is $s_0^- - (1 - R)b$, and the inclusive value is minus the log of the outside share, so the availability term per unit of Blue Bell's share is

$$\frac{1}{|\alpha| b} \log \frac{s_0^-}{s_0^- - (1 - R)b}, \quad (\text{E.5})$$

which falls as R rises for given b , s_0^- , and α , and does not depend on how the retained demand is allocated.

F The Rival Price Decline and the Retention Benchmark

The observed effect of Section 3.1 estimates the left side of equation (18), the change in rival sales between the 2015 market without Blue Bell and the market 2015 would have had with it. The evaluation design predicts the first term on the right, the effect of Blue Bell's absence with the rest of the market in its 2015 state. The second term, rival sales that the recall changed through other channels, is in the benchmark and not in the predictions. This appendix assesses the possible magnitude of one channel, the net decline in rival prices of Table 12, and asks how large it would have to be to change which model has the lowest misplaced share.

The calculation is an order-of-magnitude calibration with three inputs. Rival gallons in the scored market-months are about 5.7 times Blue Bell's 2014 gallons. The net decline in rival prices is 4.4 percent. For the elasticity of total rival demand to a uniform change in rival prices, I use an illustrative range of 0.3 to 1.4; it is a range for the calibration, not each model's computed elasticity. If the decline was a response to the withdrawal, the rival sales it produced are about 5.7×0.044 times that elasticity, between 0.07 and 0.35 gallons per 2014 Blue Bell gallon. The range is conditional on the elasticity range and on attributing the decline to the recall. Such a gain would sit in the observed retention of 0.77, and it moves the benchmark for every model alike.

For this sensitivity exercise, I allocate the hypothesized price-induced gain across destinations in proportion to their observed gains. This holds the observed allocation shares fixed while lowering retention. A gain concentrated on one destination, such as Dreyer's, whose price fell most, or private label, need not give the same thresholds. Table F.1 lowers the benchmark in this way and leaves the models' predictions as they are. A specification with PL-pooled nests has the lowest misplaced share for a benchmark above 0.591, that is, as long as the gain attributed to the price decline is below 0.183 gallons per 2014 Blue Bell gallon: the nested logit down to 0.770 and RCNL below that. Between 0.591 and 0.451 the RC model has the lowest, and below 0.451, a gain above 0.324, logit does. In the calibration above these gains correspond to elasticities of about 0.73 and 1.29. The thresholds concern which model has the lowest misplaced share; the order among the other models changes at other benchmarks. Lowering the benchmark brings the nested logits' retention to or above it and logit's and the RC models' closer to it, so the adjustment works in the direction of the non-nested models.

Whether the decline was a response to the withdrawal is left open. Three facts bear on it. The models' own pricing simulations predict a small increase in rival prices when Blue

Table F.1: Misplaced share when the benchmark is lowered proportionally across destinations

Benchmark retention	Gain removed from the benchmark	Logit	NL, PL separate	NL, PL pooled	RC	RCNL	Group RC
0.774 (observed)	0	0.350	0.291	0.101	0.322	0.104	0.340
0.70	0.07	0.279	0.249	0.112	0.256	0.075	0.273
0.65	0.12	0.234	0.221	0.141	0.211	0.113	0.228
0.62	0.15	0.207	0.204	0.164	0.184	0.135	0.201
0.60	0.17	0.189	0.212	0.179	0.166	0.150	0.183
0.55	0.22	0.144	0.235	0.217	0.123	0.188	0.138
0.50	0.27	0.100	0.259	0.254	0.087	0.235	0.100
0.45	0.32	0.064	0.284	0.293	0.064	0.285	0.064

Misplaced share of Blue Bell's volume, equation (22), with each destination's observed gain scaled by the ratio of the benchmark retention to the observed 0.774 and the models' predictions unchanged. The second column is the observed retention minus the benchmark, the rival sales attributed to the price decline, allocated across destinations in proportion to their observed gains by assumption. The lowest misplaced share changes from the PL-pooled nested logit to RCNL at a benchmark of 0.770, from RCNL to the RC model at 0.591, and from the RC model to logit at 0.451. The placebo-net benchmark of Section 8 is a different adjustment, not reproduced by this table.

Bell leaves, not a decline (Section 7.2). The decline does not increase monotonically with prior Blue Bell exposure, unlike the gradient the pricing simulations predict: it is 6.5 percent in the medium-exposure markets and 2.7 percent in the high-exposure ones (Table 12). And Dreyer's, whose price fell most among the rivals, had the gain that, of the six models, only the PL-separate nested logit reproduces (Section 6.2). Retail prices can fall when demand shifts toward a product: Chevalier et al. (2003) find that supermarket prices fall during seasonal demand peaks, mainly through retail margins, so a decline is not by itself evidence against a response to the withdrawal. The placebo-net benchmark of Section 8 is a different adjustment, for the pre-trend rather than for the price response.