Marketing Analytics, Inc.
500 Davis Street, Suite 1010
Evanston, Illinois 60201
Tel: (847) 425-9999
Fax: (847) 425-1332
www.marketinganalytics.com

November 19, 2004
Revised: 12/22/2006

Subject: **Impact of Improved FSP Data on Base Price Elasticities**

This provides Marketing Analytics' position on the increase in estimated base price elasticities that occurred with the recent release of IRI's improved FSP (Frequent Shopper Program) pricing data.

Background

This summer, IRI implemented a significant improvement in their handling of FSP pricing. IRI is now incorporating discounts available through FSP cards into all their price, dollar, and discount measures, provided the data is available from the retailer. Previously, only *advertised* FSP prices were available in the data, and they were only reported in selected, special FSP measures – not integrated into all measures. Thus, most IRI price models, including those in Drivers on Demand™ (DoD), omitted FSP pricing as the measures were incomplete and not fully integrated with the rest of IRI's data.

When price models were run on the new, improved, and fully integrated FSP data, base price elasticity estimates increased significantly – typically 40% or more on heavily promoted items. While promoted price elasticities, feature lifts, and display lifts were expected to change, the extent of the base price elasticity changes surprised many of us. This white paper explains why the elasticities changed and what, if anything, should be done about it.

Overview of Findings and Conclusions

Estimated elasticities for base price changed for the same reason they changed for promoted price, features, and displays: all these variables were correlated with FSP discounts. It's a mathematical fact that a variable omitted from a model will cause coefficient errors for a variable in the model if the two are correlated. The nature of the correlation was such that base price tended to be high when the unobserved FSP promotions were active. Without the FSP price data, the model had difficulty explaining high FSP-related sales in the face of base price increases and thus incorrectly concluded base price elasticities were low.

The upshot of this is that DoD and other IRI price model elasticities – base price, promoted price, feature, display, and other – should be correct and have been

from the first day models were run on the new data. There is no need to change modeling methodology. Managers should make pricing decisions with the knowledge that base price elasticity is higher than previously estimated when we were blind to FSP pricing.

The rest of this document discusses analysis details.

Analysis Details

Test Data

Unless noted otherwise, models and analyses for this study are based on 70 weeks of data from 3/31/2002 through 7/27/2003 for 210 Promoted Product Groups ("PPGs"), spanning 43 brands, and accounting for about 80% of the volume in the Carbonated Soft Drink (CSD) category. The 70 weeks resulted from matching observations so that only PPG-store-weeks present both before and after the data change were included. An example of a PPG is Diet Pepsi Caffeinated 12oz 12-pack cans. All 3,000 grocery stores in IRI's sample were modeled at the PPG-store-week level, providing a little over 30 million observations. From the 30 million observations, elasticities were estimated by PPG-RMA for the 210 PPGs in each of 148 RMAs (Retailer Marketing Areas). Excluding PPG-RMA combinations without distribution, we ended up studying a little over 25,000 PPG-RMA base price elasticities. These 25,000 elasticities were estimated repeatedly using a variety of price variables, base price algorithms, volume data, etc. Thus, altogether, this analysis is based on summaries of about a half a million base price elasticity estimates.

Test Model Specification

Model specifications in all cases were a slightly simplified version of the DoD store level model specification, which itself is identical to IRI's Promo Drivers model but with the addition of a linear trend term (week number).

Model Specification Used in Tests

$$\text{Log(Volume)} = \beta_0 + \beta_1 \text{Log(Base Price)} + \beta_2 \text{Log(Price / Base Price)} + \beta_3 \text{Feature 0/1} + \beta_4 \text{Display 0/1} + \beta_5 \text{Feature\&Display 0/1} + \beta_6 \text{Feature}_{t-1} \text{ 0/1} + \beta_7 \text{Display}_{t-1} \text{ 0/1} + \beta_8 \text{Feature\&Display}_{t-1} \text{ 0/1} + \beta_9 \text{SeasonalityIndex} + \beta_{10} \text{WeekNumber} + \beta_{11} \text{Holiday1 0/1} + \ldots + \beta_N \text{HolidayN 0/1}$$

- Data is at the PPG-store-week level
- Coefficients are by PPG, except Base Price (β_1) and Price (β_2), which are random effects by PPG-RMA

- For purposes of this test, no sign correction was performed
- Observations are weighted by base volume
- *Key simplifications versus full DoD model: no competitive effects, no sign correction, only random price effects, random effects by RMA instead of by store*

Elasticity Due-To Analysis

Our elasticity "due-to" analysis started with data that produced the low elasticities previously seen, and then altered one part of the data at a time until the high elasticities were produced. The change in elasticity at each step is the change "due to" that data change. At every step, the entire 30 million observation database was modeled. The procedure was performed once using IRI's base price algorithm, and once using Marketing Analytics' base price algorithm. This analysis allowed us to measure the separate contributions to the elasticity change from the following three differences between Pre-FSP and Post-FSP data (before and after the data change):

- The volume and merchandising data changed somewhat (data revisions, improved alignment of feature and display data to scan weeks)
- The price data changed a lot (now includes FSP discounts)
- The base price changed as a result of the price data changing (base price is calculated from the price data)

Thus, we first modeled the Pre-FSP data, then changed the volume and merchandising data to Post-FSP and re-modeled, then changed the price data and re-modeled, and finally we changed the base price data and re-modeled. In three steps, we "morphed" the Pre-FSP data into the Post-FSP data, measuring the elasticity at each step.

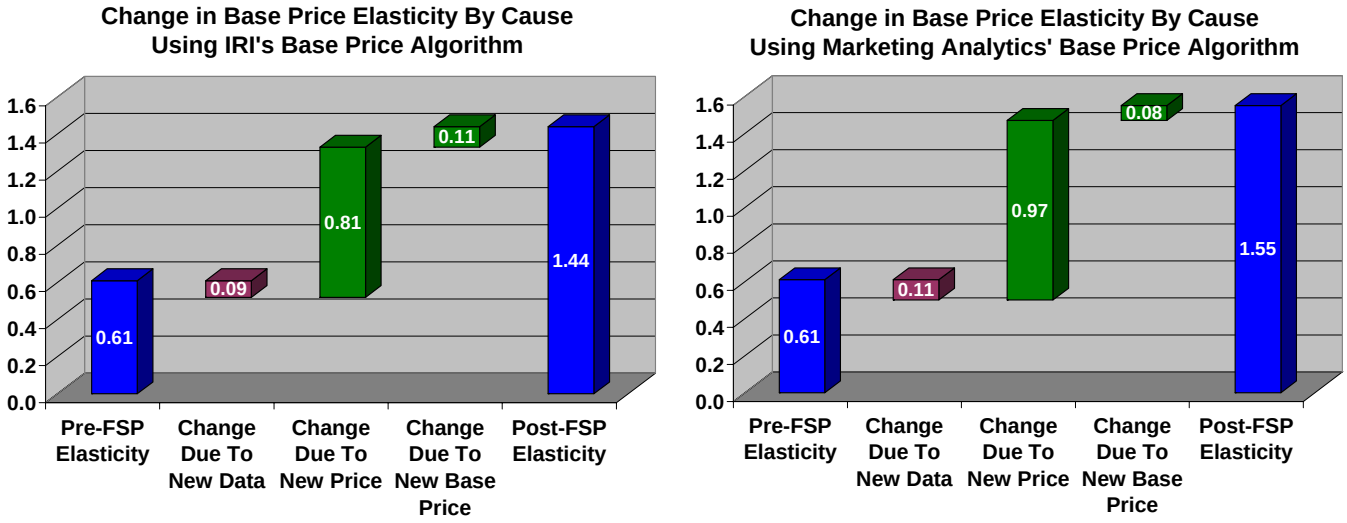
Base price elasticities were estimated by PPG-RMA and then weighted by base volume to get an overall average across all modeled PPG-RMAs in the category. Average base price elasticities from the four tests appear below.

	Average Base Price Elasticities	
	IRI Base Price Algorithm	MA Base Price Algorithm
Pre-FSP data	-0.61	-0.61
Post-FSP volume and merchandising with Pre-FSP price and base price*	-0.52	-0.50
Post-FSP volume, merchandising and price, with Pre-FSP base price*	-1.33	-1.47
Post-FSP data	-1.44	-1.55

Average across the 175 largest CSD PPGs with consistently right-signed elasticities in all tests.
** Procedurally, Pre-FSP price data was merged into the Post-FSP dataset by PPG-store-week*

Note these elasticities are lower than we are used to seeing in CSD because no sign correction has been performed. However, the increase is what we're studying and it occurs with or without sign correction.

Taking differences between the various models above, we constructed the base price elasticity change "due-to" analysis shown below.



Average across the 175 largest CSD PPGs with consistently right-signed elasticities in all tests. "New Data" above means new volume and merchandising data.

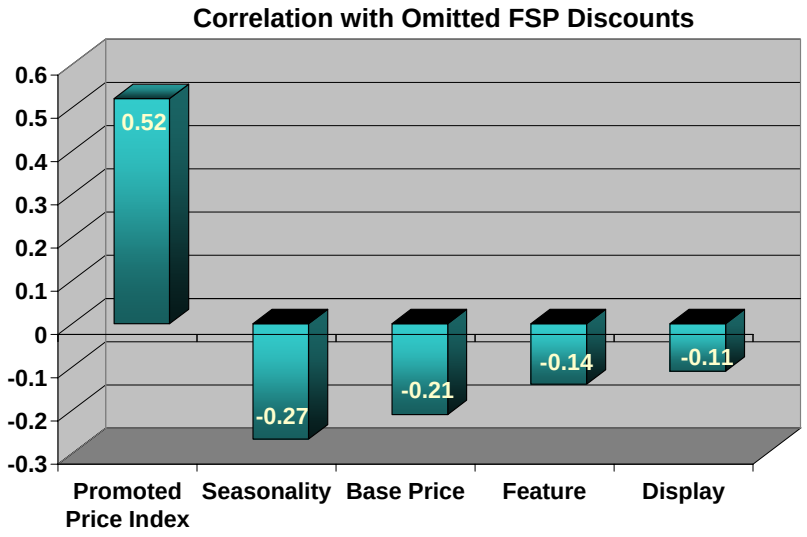
Thus, reading from the chart on the left, 0.81 or 98% of the 0.83 net change in base price elasticity (from 0.61 to 1.44) is due to the improvement in price data *separate from any impact it had on base price*. 0.11 or 13% of the net change is due to changes in base price. The same story holds if we eliminate IRI's base price algorithm and use Marketing Analytics' algorithm. *Thus, the primary reason for the increase in base price elasticity is the change in the price data itself.*

It is important to be clear on what these due-to tests prove <u>beyond any doubt</u>:

1. The change in elasticities is not due to a flaw particular to IRI's base price algorithm. The same increase is seen using Marketing Analytics' algorithm.
2. The bulk of the change in elasticities is not due to anything about base price. The exact same base price data can produce a large base price elasticity or a small one: all you need to do is change the price data
3. The change in elasticities is not due to anything about the model specification. The same specification can produce a large base price elasticity or a small one: all you need to do is change the price data

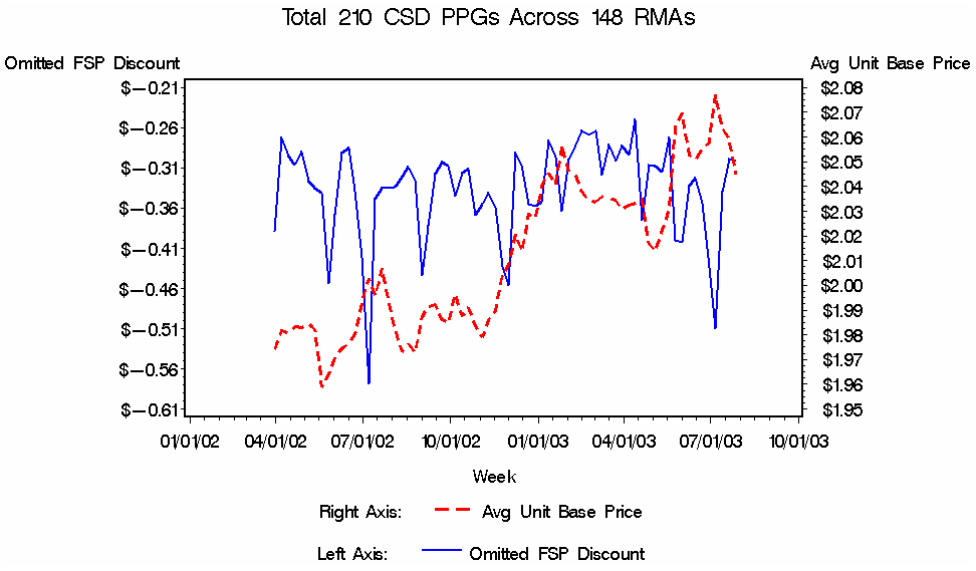
Why Price Data Impacts Base Price Elasticity

Changing or adding a variable to a model impacts the coefficient for another variable *when the two are correlated*. This is why estimates changed for promoted price, features, and displays. These are obviously correlated with FSP discounts, so adding FSP discounts to the model would naturally impact these coefficients. The surprise was that base price was also correlated with FSP discounts. But it clearly was, and even more than features and displays were.



Correlations calculated by PPG-RMA, using PPG-RMA-Week data for 200 CSD PPGs and all FSP chains, then weighted by base units, and averaged across PPG-RMAs. 10 Private Label PPGs excluded. FSP discounts correlated with seasonality as retailers promote soda in summer

It can be seen even at a high level that there is a tendency for base price and the omitted FSP discounts (Post-FSP minus Pre-FSP unit price) to mirror each other.



What Caused the Correlation?

The correlation came from an inconsistency in the handling of extended price promotions:

- During extended non-FSP price promotions, base price went down
- During extended FSP promotions, it didn't

Base price didn't go down during extended (or any) FSP promotions because price itself didn't go down.

Here's how that works when we have three possible states: regular price, normal (non-FSP) promotion, and FSP promotion.

(a) For regular price weeks, everyone pays regular price and that's the price reported to us
(b) For normal (non-FSP) promotion weeks, the retailer charges a lower price to all customers. This discount is included in our data, so we see the lower promoted price
(c) For FSP promotion weeks, the retailer charges regular price to non-FSP customers (of which there are typically very few) and a lower FSP price to FSP customers (almost everyone). But FSP discounts were omitted from the data. So what gets reported is regular price for *all customers*

This is the standard 2-tier pricing system. 3-tier systems also exist, where there are promotions with 2 discount depths: one for non-FSP customers and a deeper one for FSP customers. The situation that causes the correlation applies there as well, just in a more complicated way.

When either of the tiered systems is run with extended promoted periods, and FSP discounts are omitted, the correlation is created. The simulated example below illustrates this using a 2-tier system in the following rotation: 3 weeks at regular price, 20 weeks on normal promotion, 20 weeks on FSP promotion, then repeat the cycle. With this pattern, a -0.63 correlation is created between base price and FSP discounts. The longer the extended promotions, the higher the correlation. For very short promotions (e.g. 1 week), there is no correlation.

The example below uses IRI's base price algorithm, but very similar patterns occur with Marketing Analytics' and Nielsen's algorithms.

After the simulated example is an example using actual data from one of the largest national FSP chains. When comparing the simulated and actual examples, note that the simulated example used both extended FSP and non-FSP promotions, while the actual example has extended non-FSP but short FSP promotions. Extended non-FSP promotions are required to get the correlation: extended FSP promotions are not.

Simulated Example of Induced Correlation Between FSP Discount and Base Price

FSP-blind price goes up to regular price at FSP promo

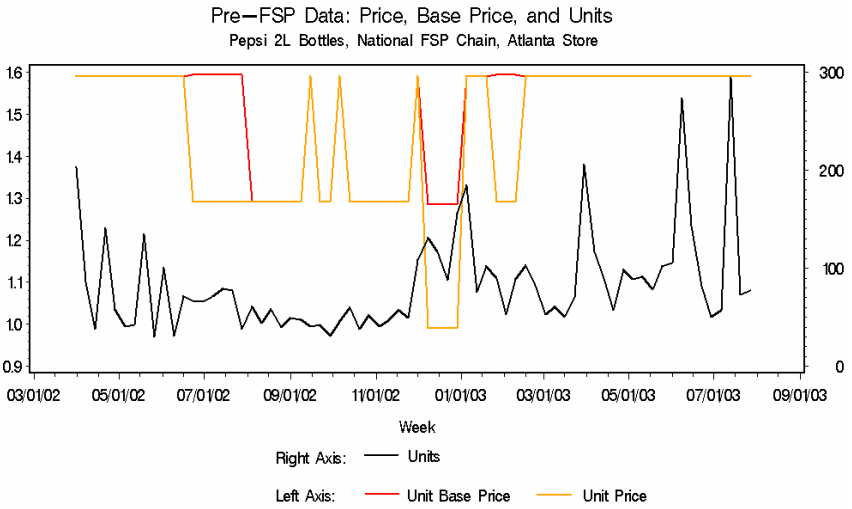
FSP-blind base price also goes up to regular price at FSP promo

FSP-aware price stays low for very extended periods

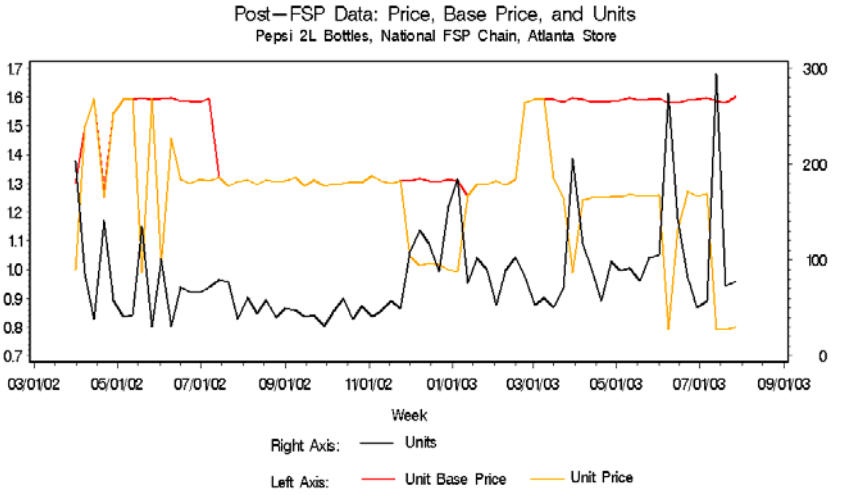
FSP-blind base price plotted next to FSP discount (Post-FSP Price - Pre-FSP Price)

-0.63 correlation

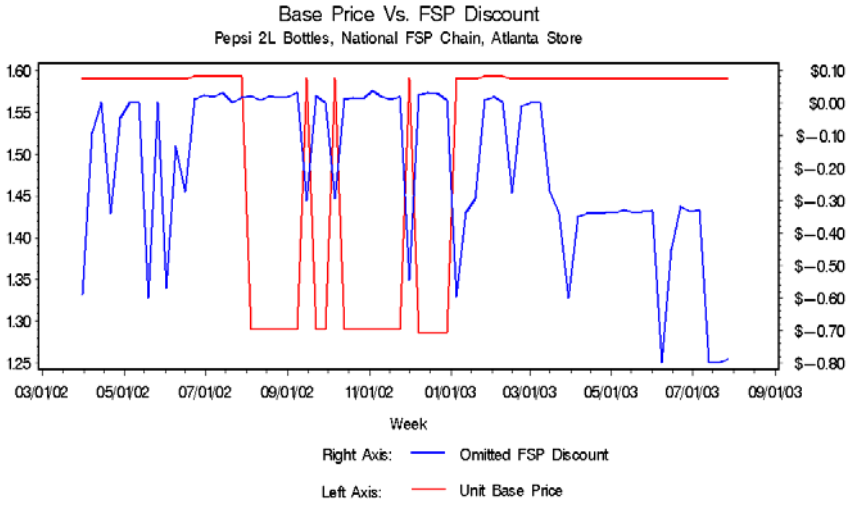
Actual Example of Induced Correlation Between FSP Discount and Base Price



In the FSP-blind data, we see an extended price promotion from July 2002 through March 2003. It appears to be interrupted by four brief regular-price periods: 3 spikes and 1 longer period.



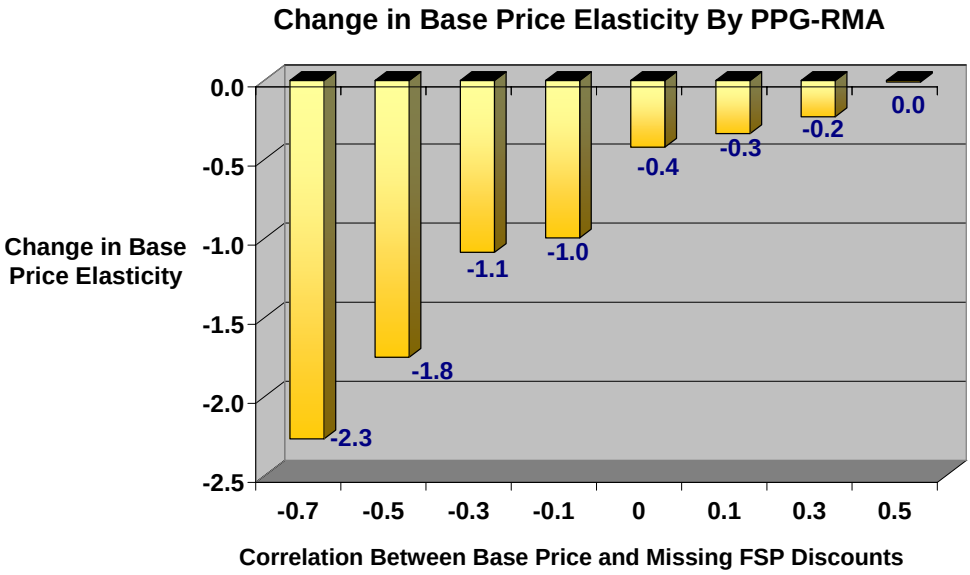
In the FSP-aware data, we see that those four periods were really four FSP promotions, so it was really one very long extended price promotion.



This shows the difference between the first two graphs – missing FSP promotions, next to the FSP-blind base price. The induced correlation between the two is clear.

Correlation Drives the Elasticity Change: Confirming Analysis

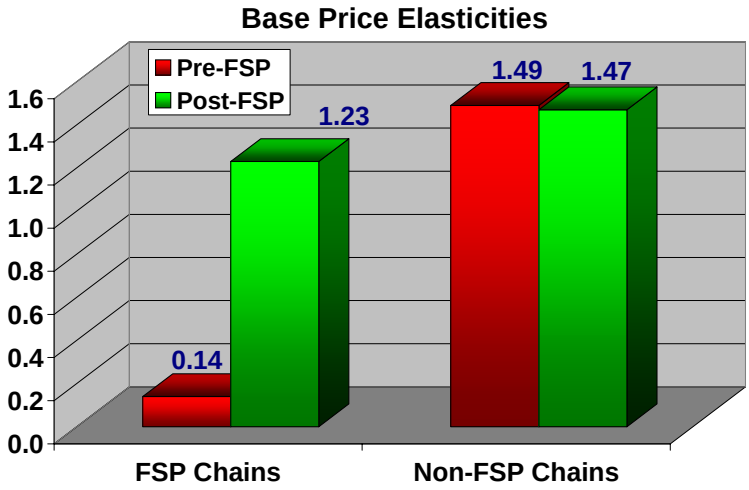
To double-check that it's really the correlation between base price and missing FSP discounts that drives the change in elasticity, we checked to be sure the largest changes in base price elasticity at the PPG-RMA level are for the PPG-RMAs with the highest correlations. They are.



Average across 200 CSD PPGs and all FSP chains. Top 8 correlation buckets only (i.e. while each bar above is based on an average of 1600 PPG-RMAs, only 29 PPG-RMAs had a correlation of 0.9 so 0.9 correlation was omitted). Note the correlations are between a variable in the model (base price) and one that is not in it (missing FSP discounts).

Elasticities Are Truly High: Confirming Analysis #1

To double-check that the right elasticities are really the higher ones, we ran the model separately for FSP and non-FSP chains, using data before and after the FSP data change. If elasticities are really, truly high, and the only thing dragging them down is a problem with the FSP data prior to the data revision, we should see all high elasticities except for FSP chains prior to the revision. We do.



Average elasticities across the 175 largest CSD PPGs with consistently right-signed elasticities in all tests

Elasticities Are Truly High: Confirming Analysis #2

As yet another confirmation that the issue is that base price elasticities were being inappropriately dragged down before the data change, we dug into the RMA with the largest elasticity change for two PPGs we were studying in depth. "Big Y Supermarket" in western Massachusetts showed a very large elasticity change for Pepsi 2 Liter Bottles. For all this to hang together, we should see that the observations that are most dragging down elasticity for this PPG-RMA tend to have this correlation problem.

Thus, we looked at the influence of each individual observation, and ranked the observations in order of their downward drag on base price elasticity. Sure enough, the #1, #2, and #3 observations dragging down base price elasticity for this PPG-RMA were all situations where:

1. Base price was higher than usual, *and*
2. Volume was very high, *and*
3. The reason volume was high was there was a particularly deep FSP promotion going on, *and*
4. That FSP promotion was missing in the Pre-FSP data

Thus, as hypothesized, the most depressed elasticity was clearly being depressed because the model needed to explain high volume from an unidentified FSP promotion in the face of high base prices.

Correlations Not Result of Base Price Spikes: Confirming Analysis

All major base price algorithms (IRI's, Nielsen's, Marketing Analytics') have provisions that prevent base price from ever being dramatically below price. Basically, they jump up to prevent this, though the algorithms vary in exactly when they jump and how long they stay there.

We know this is not having a direct, significant impact on base price elasticity because even with these jumps, we can produce high or low elasticities simply by changing the price variable. What we wanted to prove in this confirming test was that the largest of these jumps were not having a major *indirect* impact on base price elasticity by largely driving the correlation between base price and discounts.

Thus, we eliminated all observations containing jumps in base price of 20% or more, plus the following 2 observations for good measure. This deleted about 500,000 or 2% of the 30 million PPG-store-week observations. And it had virtually no impact: the low elasticities stayed low, and in fact lowered by a little less than 1%.

In summary, this analysis was just to confirm that the correlation was *not* the result of the largest jumps in base price. It's not.

Lessons Learned

The purpose of this analysis was to determine which base price elasticities are right: the old, lower ones or the new, higher ones. The answer is the new, higher ones are right. As modelers, we can be proud that our models have been and still are working well, given available data. But there is still a lesson here and a gray area worth discussing.

Lesson: Expand "Available Data"

The fact is the old base price elasticities were biased downward for highly promoted products in FSP accounts, and the root of the problem was we were missing important data: FSP discounts. It looks like that's fixed now. But the solution can be generalized to "Don't let your models run indefinitely without important sales drivers, particularly ones that could be correlated with key variables that *are* in the model".

This time, it was omitted FSP discounts. But manufacturers and retailers do other things that could be correlated as well. Advertising could accompany a package change, major promotion, or even a base price change. This experience should spur us to move DoD toward more inclusive, full marketing mix models as soon as possible, and not just to get new coefficients – but potentially for the accuracy of existing coefficients as well.

Gray Area: How to Handle Extended Price Promotions?

It is *not* true that the reason base price elasticities were depressed was because of a problem with the base price algorithm. But it sort of sounds that way in terms of the way it handles extended promotions, so an explanation is required.

We know the base price algorithm is not the problem because we can fix the elasticities (increase the old ones) by fixing the price data, regardless of the base price's behavior. Base price can always follow extended promotions, never follow them, or bounce back and forth between following and not following. In all cases, fixing the price data brings elasticities up to where they should be. The problem was the price data and the solution was to fix it.

However, if we *don't* fix the price data, the base price algorithm needs to be consistent. When there's an issue with the price data, and the base price switches methodologies in lockstep with the issue, we have a base price elasticity problem. Again, the methodology-switching behavior was induced by the missing FSP prices, so while it's true that if we fix the price data, this switching behavior doesn't really matter, as a side effect, it also gets fixed.

While we're on the topic of extended price promotions, it's reasonable to ask what is the best way to handle them going forward. We know we can handle them in a variety of ways as long as the price data is OK. We also know that

even with certain price data issues, we can handle extended promotions in either of two ways (follow or don't), and as long as we're consistent, we'll get reasonable elasticity estimates. But which is "best": following or not following?

Let's look at this from two angles: practical and theoretical. From a practical standpoint, it's often difficult or impossible to divine what the retailer's intended regular price is, given available data. Extended promotions can be 8 months long (see page 8). 4 months into it, looking only at price data, it has no end, so how can you possibly know it's temporary?

Looking to other data sources, there's no complete source of store-level regular price at this point. First, neither IRI nor ACNielsen have successfully gotten retailers to routinely provide regular price in their data feeds. This doesn't surprise us, as our experience is that retailers often store regular price in a different computer system from promoted or actual price.

Second, maybe we could get two prices from retailers, where one is the post-FSP price we just switched to and the other is the pre-FSP price we just replaced. But that would only give us true regular price during FSP promotions.

Finally, we obtained advertised regular price from Market Track Inc. to see if it could help us. In terms of getting a better understanding of what was going on broadly, this data was helpful. But it didn't purport to be and couldn't be a source of store-level regular price for all retailers. While a few retailers put regular price in their feature ads (e.g. Winn Dixie, Jewel, A&P), most do not (e.g. Kroger, Safeway). Further, given the popularity of price zones, advertised regular price is usually of necessity either a range or just one selected price zone, so it can't be reliably matched to stores.

Thus, store-level regular price is not generally available at this time. Without that, we're looking just at price and in this case, we really can't practically avoid following price down at some point during a very extended price promotion.

From a theoretical standpoint, even if we could divine retailers' regular price, we might not want to. At the insistence of a retailer who sells a high proportion of their products "on sale", Marketing Analytics used actual regular price, which they gave us, instead of base price in all our algorithms. But regular price for "always on sale" products can be close to fiction, causing model problems. For example, if you use these fictional base prices, you're always on discount, leaving promotional baselines with zero non-promoted observations – a precious commodity for them. Sales models need to be based on actual occurrences of people buying things. The farther you get from this, the more your models are based on theory – not reality. When nobody pays full price, what does it matter what full price is? Double it. Triple it. Nobody cares and sales aren't impacted.

To summarize Marketing Analytics' point of view on the debatable topic of regular price, base price, and extended promotions:

- Practically speaking, regular price is not generally available or discernible
- Even if we knew regular price, we wouldn't use it for modeling purposes
- We want to know the highest price a reasonable number of people actually paid recently, and recent discounts that have been offered
- Given volume at those prices, we can calculate a reasonable, fact-based elasticity

We realize base price is in fact partially reflecting extended promotions. We don't find that's a big problem. On the other hand, our experience is that using fictional prices in models is a serious problem. If you want to translate base prices into regular prices after the analytics are done for reporting purposes, there are ways to do that. But the models themselves should be based on as many real consumer transactions as possible.

Summary

- We've run a lot of tests. The new base price elasticities are right. The old ones were depressed for the same reason promoted elasticities, feature and display lifts were: they were correlated with FSP discounts.
- The models are fine as is. Use the new elasticities for pricing decisions.
- We should avoid this in the future by incorporating all key sales drivers wherever possible.
- We know base price dips and follows extended FSP and other promotions and we're OK with that, but we understand this is debatable

I personally want to thank Larry Menke of Marketing Analytics Inc. for his hard work and insights on this issue. Larry put a huge amount of time into this and designed the analysis that swapped price variables between the datasets, providing the most solid proof of what the real issue was. Thanks also to Bruce Richardson and Avu Sankaralingam of IRI for their insights and the extensive testing they did of IRI's existing base price algorithm and modified versions of the algorithm that they designed for this analysis. Thanks to Market Track Inc. for providing advertised regular price. Finally, thanks to many others too numerous to mention who contributed to resolving this issue.

Ross-boy Link
Marketing Analytics Inc.