Retailer-Level Modeling vs. Store-Level Modeling – Pros & Cons

In the 30+ years that scanning data has been a staple of CPG analysis there has been an ongoing debate about how to best model the data. In the late 1980s and early 1990s IRI and Nielsen had almost exclusive access to this data and modeled it at the store/week-level. But with an explosion in the number of retailers available on Nielsen and IRI databases and the ubiquity of those databases, manufacturers and other consulting companies have been modeling the data at the retailer/week-level.

POS data providers (e.g. Nielsen or IRI) have contended that modeling at the store/week-level is an advantage. Conversely, over the last 20+ years other research vendors and consultancies have pioneered approaches to model the data from the databases at the retailer/week-level. They argue that these approaches yield comparable or better results do store-level models and also have some additional advantages. So – what are the points for and against each viewpoint?

The **arguments in favor of store/week-level modeling** are generally academic in nature in that they revolve around classical statistical concepts:

- **Sample-size**: If the given retail banner has 100 stores and 20 syndicated sample stores, models using store/week data from the sample stores have 20 times as many data points vs. models using retailer/week data. More data provides tighter estimates (higher confidence intervals).
  - ✓ Note: all other things being equal, greater sample size in itself does not drive different results – it just drives tighter confidence intervals.
- **Aggregation bias:** Aggregating store/weeks into retailer/weeks, by definition, will result in data being averaged across stores for each week. In some cases this can be problematic when the variables being modeled lose some of their variability due to averaging:
  - ✓ At higher-levels of aggregation (such as market-level modeling) the stores being aggregated have a wide variability in price and promotion. For instance, at the Total Chicago level we are averaging prices at Jewel, Mariano's, Meijer, etc.
  - ✓ At the retailer-level there is much less variability in pricing and promotion across stores and therefore much less aggregation bias:
    I. An ad that is run across all of the stores in a retailer with the same price discount will have no less variability at the retailer-level than at the store-level – no aggregation bias.
    II. Some retailers have price-zones – that is, some stores will be priced differently than others. So at the

retailer-level we would be looking at average base price, and to the degree that base prices vary across stores inside the retailer there would be some level of averaging the absolute base price.

III. However base price changes are typically taken across all stores in a retailer, even if from slightly different starting base price points. In this case there would be minimal reduction in variability when looking at base price changes at the retailer-level vs. in individual sample stores.

The **arguments in favor of retailer/week-level modeling** are generally more practical in nature and take a broader perspective. The debate is not simply between store/week modeling and retailer/week modeling. These are just two choices along a spectrum of complexity.

At one end of the spectrum we could try to model a very minimal data set (say, 52 weeks of aggregated Total US data from the database).

- This would be very fast, very inexpensive and we would be very dubious of the results.

At the other end of the spectrum we could try to model with every available data input available at the most granular level possible. In this case we might include store/hour data for all stores in the US, individual household purchase data for all consumers, TV and radio TRP's by the hour, all internet advertising data, FSI drops by week and by newspaper, weather variables etc, etc, etc.

- In theory this approach could result in more accurate predictions
- But costs would be prohibitive and the project wouldn't be finished in time to be useful.

The real question is ***what is the right balance***? Proponents of retailer/week data modeling argue that some slightly larger confidence intervals are a small price to pay for dramatic savings of cost and time. Below is a quote from Doug Brooks of MMA (formerly IRI) circa 2009 from an IRI publication called "Success and Failures in Marketing Mix Modeling":

*"The first challenge is the data itself. Statisticians must accept that data will never be perfect or complete—there will always be missing data, imperfect data and/or holes in the data. Marketers and statisticians have to pick a point where the data is "good enough," selecting a point that provides the right level of insight without over-architecting the model and creatively searching for proxies to fill in the holes where needed. …. The second challenge is to recognize that it is not possible for the model to explain or predict 100 percent of sales activities. Here is where the adage, "**the perfect is the enemy of the good**" comes into play. Many companies fall into the trap of an endless and fruitless search for 'false precision'."*

Sample size will have some effect on confidence intervals, but any effect on accuracy is minimized when modeling homogeneous retailer-level data. For example, both approaches might yield -1.60 as a regular price elasticity but the store-level result could be +/- 0.05 at a 90% confidence level while

the aggregate model result might be +/- 0.10 at a 90% confidence level.

Aggregation bias is largely avoided by modeling retailer/weeks instead of market/weeks. As discussed above, promotions and changes in everyday price at the retailer are typically run homogenously across the stores in a given retailer. That is, each store in a given retailer typically does the same thing at the same time. In these cases (the vast majority of cases) there is little to no loss of variability due to aggregation bias when using retailer-weeks.

For trade promotion modeling vs. base price modeling, Nielsen and IRI databases actually provide ***store-clusters*** inside each retail banner for each week for each promotion type. For any item in a given week in a given retailer, if different stores ran different types of promotion, these store clusters are available to be modeled separately. So, we are actually modeling below the retailer-week level – each retailer week is split into five different store-clusters (non-promoted, TPR Only, Ad Only, Display Only, Ad with Display). This ensures data homogeneity and eliminates almost all chances for aggregation bias.

Finally – all modeling approaches are not the same. We believe mLogic's overall approach is superior to those of store-level modeling vendors as well as other aggregate-level modeling vendors.

Vendors who try to simply mimic the store-level modeling approach argue that their approach is "close enough" given minimal aggregation bias and is well worth the savings in cost and time. We might agree with them that the savings they drive are worth the minimal loss in accuracy – it depends on the vendor.

But we adamantly do not believe in "good enough" so we do not simply try to mimic store-level modeling approaches. As outlined in our article "Traditional Modeling Limitations & The mLogic Solution" we outline how we go further to enhance the overall system by addressing the shortcomings in the typical bottom-up PPG modeling approach.

- By including a top-down category perspective we ensure that simulations input at the PPG-level will result in more accurate at the category, manufacturer and segment levels.
- Not only does mLogic model base price, trade promotion and distribution, we also model pack-size elasticity. Pack-size reductions are perhaps the most frequent method used by manufacturers to increase price. mLogic's system ensures diminishing elasticity and incrementality as we move up from retailers to channels and then up to Total Country. This implicitly accounts for cross-retailer and cross-channel switching.
- mLogic's financial analysis system is much more accurate and complete for several reasons, especially because of our greater accuracy at higher levels of aggregation

(manufacturer, category, total channel) – our approach fully captures cannibalization!

- mLogic cleans weekly retailer data to better reflect base prices and baselines.
  - ✓ For instance, a 13-week rollback in Walmart is typically (wrongly) broken into a 6-week TPR and 7 weeks of a new base price. mLogic fixes this problem.
  - ✓ This is important not only for elasticity accuracy but also for promotion ROI analysis (using truer overall incremental volume) and for simulation purposes (user sees truer picture of promotion frequencies, mix and discounts).
- mLogic models diminishing returns in impacts for increased promotion frequency, distribution gains and allows for an S-curve in how deeper and deeper promoted discounts may drive smaller and smaller incremental unit lifts.

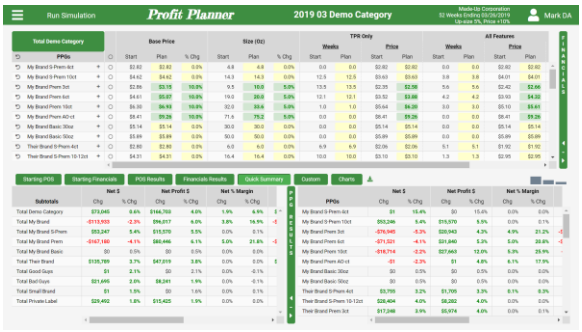
So yes – if cost and time were of no importance and <u>if you ran the exact same model</u> – once using store-level data and again using retailer-level data, store-level model results *might* be slightly more accurate (or at least you'd have slightly tighter confidence intervals).

But at what cost? The savings in time and money are very important considerations.

And mLogic does **<u>not</u>** run the exact same model – our comprehensive approach addresses multiple shortcomings of the typical store-level model and yields model results that are <u>more accurate and predictive overall</u>. mLogic provides much more complete modeling output, including pack-size elasticity, category effects, diminishing returns, etc.

mLogic's overall approach is superior in ways that count in the real world:

- More accurate model due to entire category approach, minimized aggregation bias, diminishing returns, cleaner data, etc.
- Superior and more complete modeling output, including pack-size elasticities
- Complete financial results for all players, including the retailer
- Accurate roll-ups for all levels
- Savings in cost and time
- Superior simulation tool



For more information contact mLogic Consulting at info@mlogicconsulting.com or 312.208.9675.

mLogic Consulting Inc.