# 'The Art of the Ask' in Forecast Modeling:

## Implications of "Allocation" vs. "Discrete Choice" Projections

*By Michael Polster, Ph.D. and Susan Schwartz McDonald, Ph.D.*

Growing focus on the cognitive processes surrounding customer choice has invited greater attention to the way the choice process is simulated and measured in market research. One of the greatest challenges of market research is, in fact, deciding precisely how we ask respondents to predict their own behavior, given that people are fundamentally poor prognosticators, and research artifact introduces other sources of error.

In general, sound research requires that the questions we pose to customers be sensible and relevant to the decision environment we wish to evoke. In forecasting research, we often have the option of relying on either (a) a *discrete choice task*, in which respondents are presented with a list of products and asked to choose which *one* they are most likely to purchase, (b) a *constant sum exercise*, in which respondents allocate future purchases across a range of options, or (c) *an estimated probability or likelihood* of selecting one or more offerings (via ratings or percentages).

Although each approach has its partisans (and models can be built to accommodate them all), most researchers would concede that some situations make one or the other approach easier to pursue and sometimes potentially more appropriate. What has long been missing from the debate about "best practice" is reasoned discussion concerning the sources of error associated with each metric and empirical evidence regarding the modeling implications of using one or another. Conversation on this issue has long been confounded by the fact that some research buyers use the term "discrete choice" mistakenly to refer to any sort of conjoint exercise. For the record, discrete choice and conjoint are not, in fact, synonymous; discrete choice is just one of several kinds of choice model inputs.

***Based on a systematic review of the options, and on analyses we've conducted to compare the results achieved with different approaches, we think that there is an evidence-based case to be made in favor of constant sum exercises in many routine prediction exercises where discrete choice might have been the lead contender.*** We can't prove that allocations invariably improve predictive accuracy but we can point to features and outcomes that might suggest that possibility.

### *The Natural Ecology of Real-world Decision Environments*

It is clear that real-world decisions come in a variety of types. There are frequent, often reflexive, choices (what you'll order for lunch); frequent but thoughtful choices (what a physician will prescribe for the next patient); and infrequent, often higher-order, high-stakes choices (e.g., what

car you'll drive). It's logical to be mindful of the nature of the decision when structuring a dependent variable in order to put respondents in the right frame of mind and "mood" when they decide for us.

For instance, in modeling more frequent purchase or preference decisions, questions need to give scope to all the available options by asking for allocations especially if situational variables may influence outcomes. Ordering lunch is a perfect example of that type of decision environment. It would be senseless to require respondents to pick a tuna sandwich over ham and cheese in a winner-take-all prediction of future lunches.

By contrast, when modeling durable purchases like cars, there is often a strong presumption in favor of a discrete choice based on the inference that voting for a single option rather than allocating "chits" is somehow more realistic, and therefore more predictive. *Indeed, what we have seen over the last 10 years is a growing enthusiasm for using discrete choice, even for frequent decisions.* The argument offered is that, in real life, people typically choose rather than allocate. Shouldn't market research, therefore, both acknowledge and mimic that real world decision "ecology" by asking consumers to make discrete choices all the time? Actually, the answer may be no.

### *Reality vs. Artifact*

It is important to remind ourselves that although market research can be naturalistic, there is nothing truly "natural" about it. The best it can do is evoke and approximate actual decisions – or more accurately, accumulate data that help *us* simulate real world choice outcomes. Thus, the case can be made that our attempts to "mimic" market ecology in posing research questions is more academic than real. When we give respondents a series of automotive market scenarios in a conjoint exercise and require them to "purchase" or prefer only one car, we can't legitimately persuade ourselves that this *faux* choice replicates an actual one. These exercises are always a kind of intellectual fiction or role-play that allows respondents to expose some (not all) of their decision-making proclivities. It's informative but scrubbed of real-world experience and emotion, and very much prone to over-statement of various kinds.

One decision environment in which we often see both allocations and discrete choice tasks utilized with some frequency is physician prescribing on behalf of patients. Because each patient may get a different drug, it is certainly reasonable to ask physicians to estimate the proportion who will get one therapy or another. On the other hand, proponents of discrete choice argue that physicians make better predictions when asked to anticipate what they will do for their "next" patient than when they make sweeping, aerial-view allocations across the entire population. In order for a discrete choice task to be sensible in this environment, we must incorporate specific patient characteristics into the scenario so as to supply context. The presumption is that an accurate forecast can be developed for the entire population or universe of decision units by aggregating discrete choices made across a range of scenarios. What makes this approach viable at all is that unlike tomorrow's lunch choice, there are a number of known decision drivers that encourage choice of one therapy for each patient, and probability of occurrence for certain drivers can be estimated. So it can be done, but is it better? We think not.

### *Drawbacks of Discrete Choice*

Although the rationale for discrete choice on specific buying "occasions" may be seductive, there are several drawbacks to the approach. In fact, we believe that even infrequent decisions are not necessarily best modeled via discrete choice questions.

- ***Even in rational decision-making environments, it is generally difficult to supply respondents with all the information they need to make a single choice with conviction***. Clearly, we can't meaningfully frame tomorrow's lunch decision by attaching all the relevant situational variables since whim and randomness may ultimately rule the day. But even in environments like a physician's office or a procurement department, there are many known decision influences that can't be incorporated in our representation of the "next" decision scenario. In other words, while the task may be structured realistically, missing content and context still introduce very significant sources of prediction error.

- ***Although certain types of decision situations occur with known frequency, discrete choices can be difficult to roll up into a coherent, projectable forecast***. A forecast developed by aggregating multiple scenarios needs to take accurate account of the frequency of each one in order to put "Humpty" back together again – something that becomes harder and harder to do as more contextual information is supplied to the respondent. A marketplace deconstructed into dozens of decision types or scenarios can be difficult to reassemble without introducing new types of error whose impact is not subject to measurement or correction.

- ***Empirical analysis suggests that discrete choice-based models tend to produce more extreme estimates — presumably because the discrete choice task does not allow respondents to "split their vote".*** It secures only one data point per decision scenario (i.e., the preferred product) without offering information about strength of preference relative to the second and third-runner up. The price to be paid for additional data on "also-rans" is, of course, increased survey time and respondent burden but that may be preferable to a weaker model. In a recent meta-analysis, we used share allocation data from several studies to derive discrete choice data by assuming that the product allocated the largest share would have been the one selected in the discrete choice exercise. The aggregate shares produced by each approach suggest that discrete choice exercises may overstate the ultimate popularity of a product that is widely preferred, and understate the market share of a product destined to be a marginal player.

  For example, a 20% preference share in an allocation task generally nets more than 30% preference share when the data are transformed to discrete choice. By contrast, 5% in the share allocation task often nets 1% in discrete choice. The discrete choice data also have larger standard errors than the share allocation tasks, even when the means are similar. Although it might be hypothesized that the discrete choice estimates are more accurate of actual behavior, we are unaware of any empirical evidence to support that assertion. Instead, the fact that new products are virtually always subject to a discount for overstatement suggests that the allocation data more closely approximate ultimate market share – at least for popular products.

### *A Reasonable Alternative: Allocation of Probabilities*

Needless to say, we don't ever want to pose predictive modeling questions that require consumers to make fatuous and implausible judgments because that makes the role-play of research even more tenuous. We don't, for instance, want to ask consumers to tell us how they would allocate their next 10 car purchases across brands. *We do, however, have the option in just about every market environment of asking respondents to allocate probabilities* – a concept that respondents readily understand, and that experience suggests they can easily apply. At any given moment in a real-world decision process – for instance, during the run-up to a car purchase decision – people are able to informally calculate the evolving probabilities that they will choose one option in their consideration set over another. That task gives us the best of both worlds; it evokes a decision mind frame and it provides more data reducing model instability.

### *The Bottom Line*

Constant sum exercises not only require far fewer assumptions than discrete choice tasks; they also offer more data and produce more stable models, and may therefore, be quite literally, the "dependent variable of choice" when conducting forecasting research. In cases where judgments are made infrequently, and a traditional chit allocation seems inappropriate, we recommend a constant sum allocation of *probabilities* (such that the probability of all options adds to 100% likelihood) to allow respondents to hedge their bets. If a discrete choice approach *is* used – whether for frequent or infrequent decisions – modelers and marketers may wish to consider more aggressive adjustments for overstatement (and understatement) around the most and least popular contenders.