Top Box Associates, LLC

consensus point

Prediction Market Based Volume Forecasting

Presented: March 29, 2016
CASRO Digital Conference
Austin, TX

Authors: Brad Marsh, **Consensus Point**
Ken Sobel and George Ciardi, **Top Box Associates**

Background/Objectives
Consensus Point, a Prediction Market Research company, and Top Box Associates, a new product volumetric forecasting company, entered into a joint R&D effort in 2015 aimed at deriving a valid way of using Prediction Market metrics to be used as inputs for new product sales volume forecasting.

While Prediction and Preference Market research has been shown to be very accurate in predicting market, category, and industry trends as well as future consumer behavior, little validation work has been done to evaluate this data as an input into traditional volumetric models designed to forecast the in-market future sales of new products. The purpose of the R&D was to understand if/how Prediction Market research could be used directly or adapted in order to perform volumetric forecasting.

Other Relevant Research
Although prediction markets were introduced as a new methodology to the market research industry in or around 2008, they have been considered an "emerging" method over the last several years. While they have gained significant traction as a Concept Testing method for brand clients seeking a "new" and "different" method, little actual R&D and validation has been done by the suppliers (other than Consensus Point and Top Box Associates) to validate results vs. traditional methods, based on an ability to forecast in-market success. Consensus Point has invested significant energy to remedy this lack of validation over the past few years by quantifying accuracy rates on consumer and market trends predictions as well as "soft" validation of "post-launch success" from clients. We believe this joint R&D with Top Box Associates to be the first direct effort related specifically to Fourt-Woodlock volumetric forecasting.

Preliminary Research & Methodology
Because we envisioned the application of the research as a commercial research solution offering, we also designed the research to answer a secondary question: if prediction market (PM) metrics cannot be used as simulated test market (STM inputs), can we field and collect both types of measures in the same research instrument?

In order to explore this possibility, several concepts that had been pre-tested with traditional monadic online concept testing surveys were re-tested with a sample of online panelists using the Consensus Point's Huunu™ prediction market platform and algorithm as well as traditional monadic concept measures.

We designed two primary research protocols. The first entailed collecting the traditional STM measures of purchase intent, value and uniqueness first, followed by the collection of prediction market metrics. The alternative reversed the order. See Figure 1.

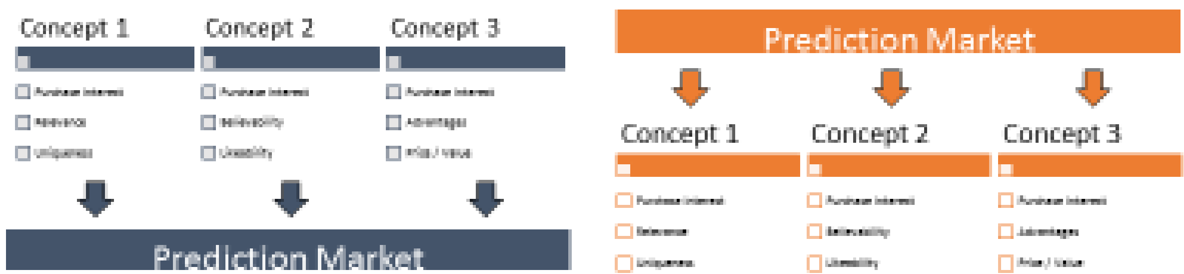


Figure 1

By having the concepts tested in two different orders, cross comparisons of the methodology can be evaluated. The key questions the design intended answer were:

1) Do consumers respond to the traditional research questions differently when asked before or after a prediction market?
2) Are the prediction market outcomes influenced by the concept exposure being shown first?
3) Could prediction market results be calibrated to serve as the consumer preference input into a traditional trial and repeat volume model?
4) Assuming #3 is possible, could results be modeled to be as or more accurate as traditional concept testing inputs to volumetric models?

An N of 1200 consumers were included in the research, representing a general population sample of consumers in the US market. That sample size provided n=200 evaluations for each concept and n=600 consumers in each of two prediction markets. The sample was well balanced across gender, age, and category incidence.

Preliminary Findings

Interestingly, our first key finding in both cases was that the metrics of the procedure in second position is biased by the preceding methodology. In other words, if data for both techniques are collected in the same survey instrument, the second set of data will be compromised. In particular, while either approach in first position can differentiate between weak and strong concepts, the data from both approaches in second position exhibits much reduced variability, which means that the ability to read good concepts from bad is relatively poor.

Further, PM metrics are very high generally when the prediction being made is not bound by time or magnitude constraints. The PM metrics magnify small differences between concepts, which helps in revealing the ranking of concepts, but can also exaggerate the absolute strength of those concepts.

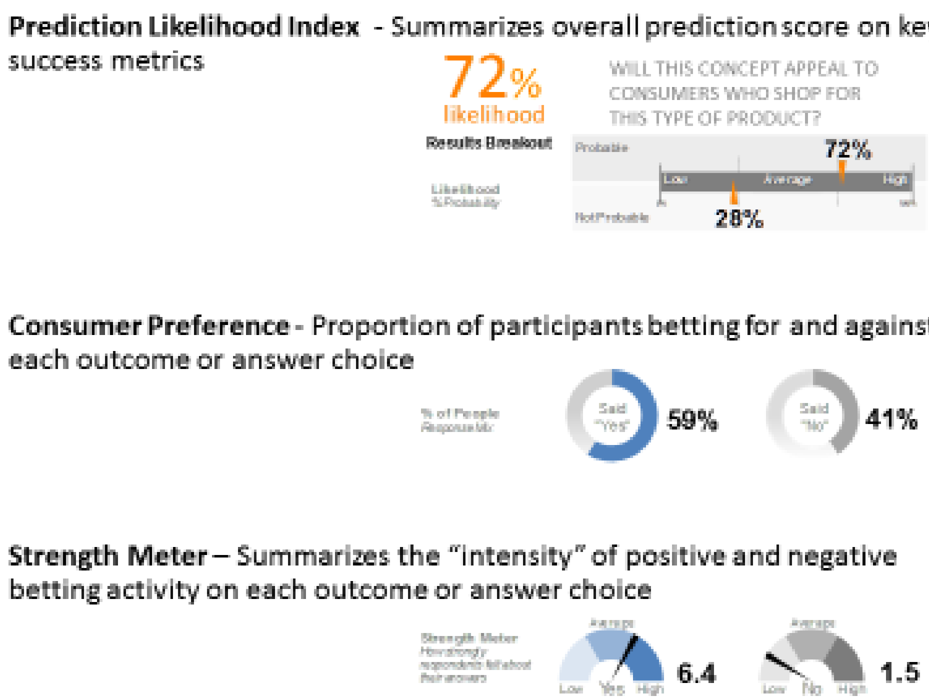
In-Depth Research on PM inputs into STM analyses

Once we determined that the "easy" fallback approach of collecting both types of metrics in the same survey instrument is not feasible, the next step was to explore whether PM metrics could usefully serve as STM inputs. (We concluded, without research, that STM metrics could not serve as inputs to a PM because STM inputs have no provision for allowing respondents to be influenced by other respondents' answers, a fundamental part of prediction markets.)

There are three primary candidate PM metrics for testing (see Figure 2):

- <u>Prediction Likelihood Index</u>: Summarizes overall prediction score (probability of occurrence) for each question asked in the market

- <u>Consumer Preference</u>: The proportion of participants betting for and against each outcome or answer choice in the market



- <u>Strength Meter</u>: Summary of the "intensity" of positive and negative betting activity on each outcome or answer choice in the market

Figure 2

Before we can perform the analyses to enable us to choose the best metric, we must also specify the target volumetric measure we seek to predict. The choices present themselves as a logical sequence:

1. **Total sales volume vs. the components of volume (i.e., trial and repeat)** – we focus on Volume Components because Total Volume depends on too many exogenous variables to be captured by any primary research (Prediction Market or traditional concept test), such as retail distribution.

2. **Trial volume component cs. Repeat volume component** – we focus on Trial Volume components because Repeat Volume components would depend on product performance and satisfaction, which requires product testing, not just concept testing.

3. **In-market penetration vs. theoretic trial** – we focus on Theoretic Trial because In-Market Penetration depends on many exogenous variables, such as distribution, couponing activity, and advertising weight.

The validation experiments showed that of the three PM metrics, the Prediction Likelihood Index is clearly the best candidate for serving as the key input used for forecasting theoretic trial. It was, however, not perfect, showing a very high degree of overstatement. Further the Strength Meter also showed promise, and we are pursuing additional research to establish whether the overall results can be improved with a composite input metric.

Research on Prediction Market starting point

An important part of the PM technique is a "stock market" feedback. Respondents are clued into the opinions of other respondents because the overall "price" of a concept is influenced by the number of respondents who "bet" on it and the magnitude of virtual currency ("tokens") that are placed at risk. The final equilibrium price of the concept is a key component of both the Prediction Likelihood Index and the Strength Meter metric. This final equilibrium price is sensitive to the **starting point** presented to respondents prior to the initial bets / responses. Customarily, the starting point is specified to be 50 percent on both the "Yes" and "No" sides of the outcome possibilities. If the "Yes" outcome is bet upon by lots of PM participants and if they collectively put many tokens at risk, these bets move the concepts value far above 50 percent. Conversely, if the "No" outcome is bet upon by lots of PM participants and if they collectively put many tokens at risk, these bets will move the market price of the concept lower than 50 percent.

In a sense, the starting point serves as an indicator to respondents about the current "crowd" prediction on a concept, and allows their judgments to nudge the value of a concept up and down accordingly. If the issue being examined by the PM is a "yes/no"

question, the use of a 50 percent start point is probably appropriate. On the other hand, if the PM is designed to address a "how much" question (as is the case for volumetric forecasting for a new product), the 50 Percent start point is probably too high for most new products, since theoretic trial rates are nearly always much lower than 50 percent.
To investigate whether a better choice of PM starting point can improve our forecast, we analyzed PMs with starting points of 50, 20, 15, 10, 5, and 1 percent.

The results were dramatic, as illustrated by Figure 3:

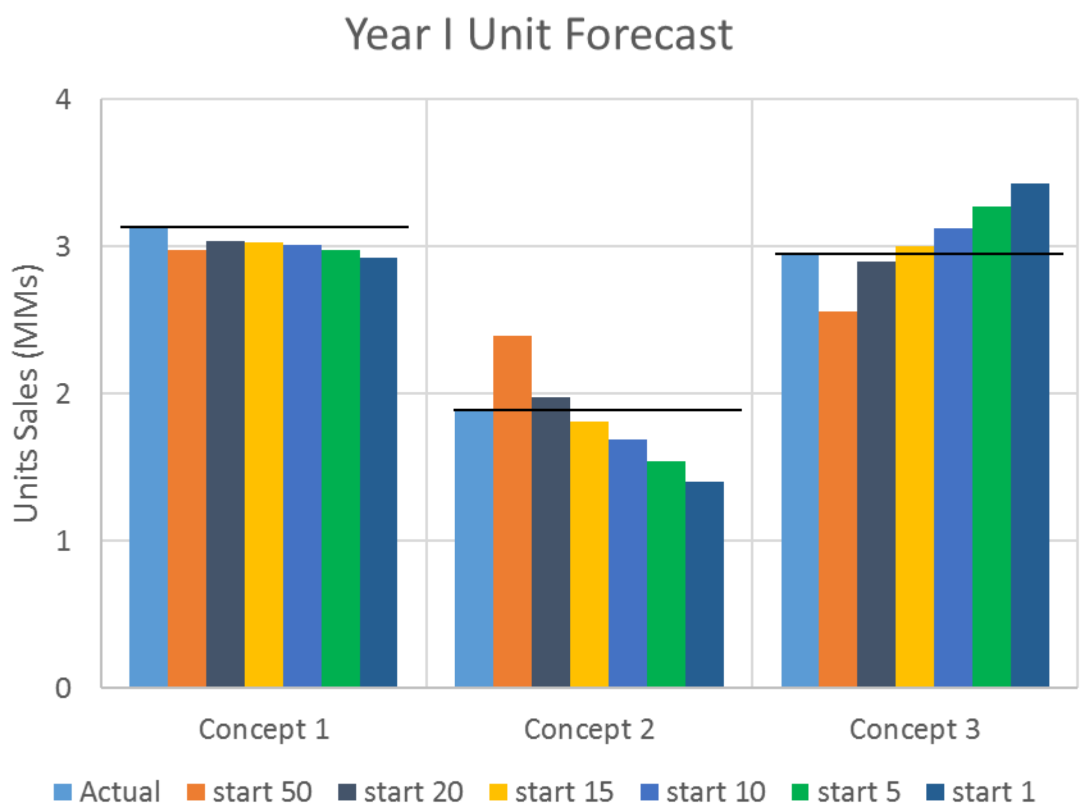


Figure 3

As you can see in the above graph, the volume estimates for Concept 1 are not very sensitive to the choice of the PM starting point, but the volume estimates for Concepts 2 and 3 are quite sensitive to the selection. For example, with the usual starting point of 50 percent, the volume estimated for Concept 2 is substantially too high, while the volume forecasted for Concept 3 is too low. If the starting point is set at 5 percent, or 1 percent, then the forecasted volume for Concept 2 is too low while the volume forecasted for Concept 3 is too high. The sweet spot for the starting point for these three concepts appeared to be in the range of 10-20 percent. At that level, the forecast for all three concepts showed very low levels of error.

As striking as this finding is, we also realized that it is not generally useful. While it works for these 3 concepts, there is no assurance that it would work for any concept. Further, we needed to know the actual in-market volume of each concept in order to select the best PM starting point. This was available for these experiments, since we selected concepts with known volumes, but this would not be generally available if we were assigned the task of forecasting the in-market sales of a new concept (which, by definition, has not yet been launched and therefore has no in-market volume).

We realized that we would need to design a procedure to select the appropriate PM start point automatically. This would have the effect of informing the respondents about the appropriate "ballpark" strength of any concept being evaluated. If the new concept was for a Ready-to-Eat (RTE) cereal, with a large number of competitors already on the market, we knew the PM starting point needed to be very small. In contrast, if the new concept was for a new brand of ketchup, with only 1 or 2 key competitors, the PM start point needed to be larger. This thinking led us to experiment with an "Order-of-Entry" (OOE) model, which predicts "par share" for the next entry in a category based on a small number of variables (number of competitive entries, time since the last entry, price of the new entry relative to the market leader, quality of the new entry relative to the market leader, and marketing support for the new entry relative to the market leader). The results of the OOE model, holding all variables equal except number of entries in the market, is shown in Figure 4:

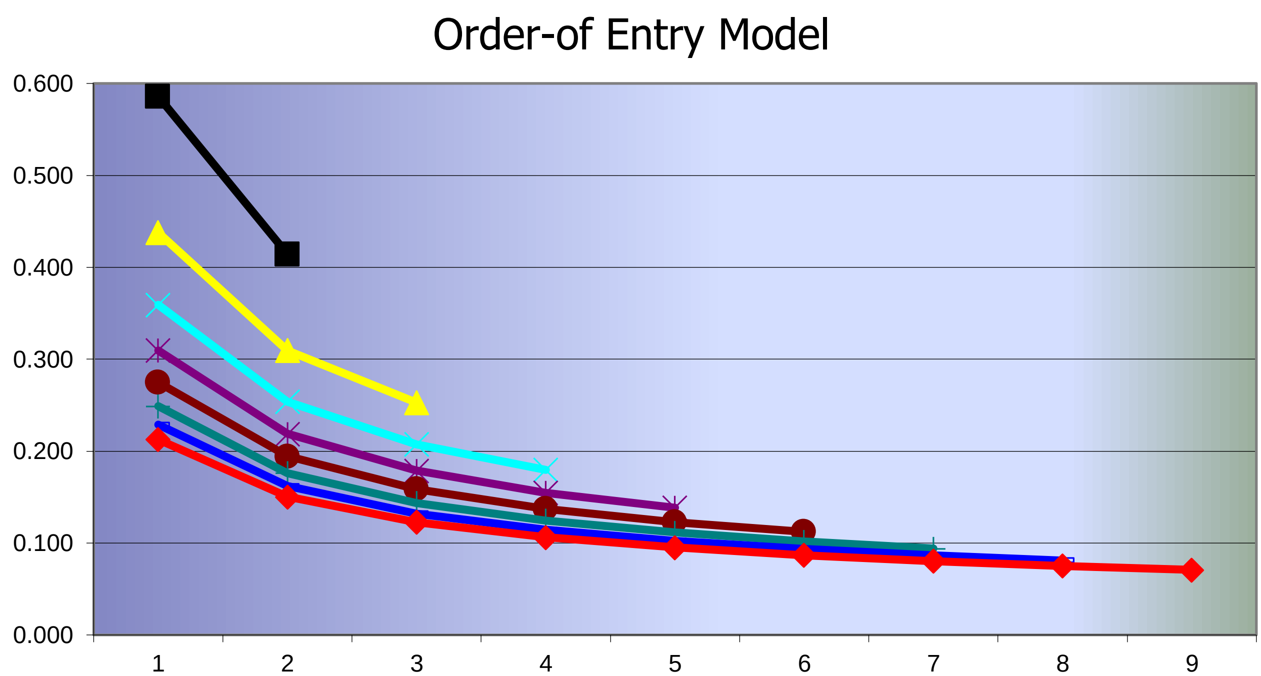


Figure 4

This graph shows that the third entry into a category with two prior competitors can expect a "par" share of about 25 percent. The ninth entry into a category with 8 prior competitors could expect a "par" share of about 7 percent. For the three concepts in our PM-STM experiment, the PM start point recommended by the OOE model is 17 percent.

By applying this OOE result to set the PM starting point at 17 percent, we were able to show forecasts for the three experiment concepts that accurately provided both relative rankings and absolute volumes of all three concepts, as shown in Figure 5.

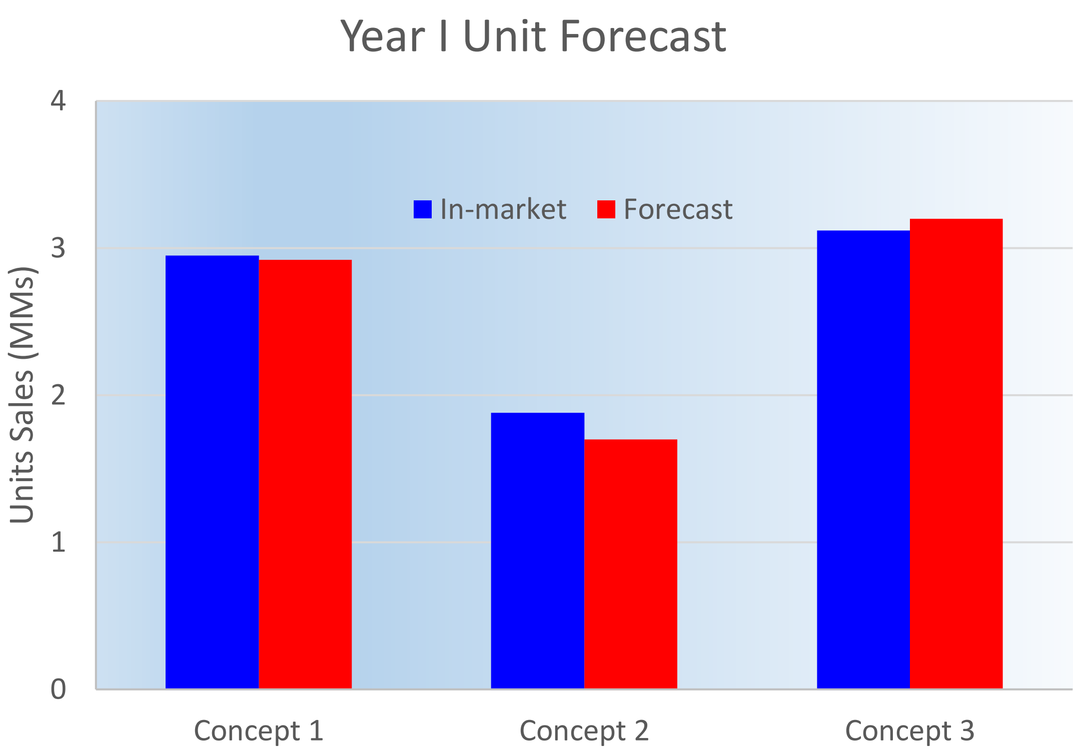


Figure 5

Summary of results
This ability to accurately forecast both rank order and absolute in-market strength results from the use of a carefully executed Prediction Market (PM) using a starting point generated by a validated order-of-entry (OOE) model and used as input into an appropriately calibrated simulated test market (STM) model informed by the marketing support used to launch the new product.

Because of this integration of models and procedures, the industry can now enjoy the benefits of Prediction Markets without sacrificing the financially relevant information from well-validated and proven Simulated Test Markets.

R&D Completion Period

September-November of 2015, results published for CASRO Digital in March 2016. Patent-pending on methodology, market game user-interface, algorithm sensitivity settings, and calibration process specific to volume forecasting.