Pros & Cons of Market Simulations Using RFC vs. HB Draws

Bryan Orme | President | Sawtooth Software

Agenda

- Review Red-Bus/Blue-Bus
- Review RFC
- Review HB Draws
- Practical Comparisons
- Conclusions & Recommendations

Market Simulators

- Use conjoint utilities to predict what respondents would choose in marketplaces made up of multiple products

- Respondents “vote” on preferred products
  - Splitting respondent votes probabilistically is more true to the logit models (e.g., HB-MNL) that (typically) were used to estimate the utilities

IIA (Red-Bus/Blue Bus)

- In the 1970s, early choice modelers were running logit models on peoples' choices of transportation within cities
- Logit models have a known property called IIA (Independence of Irrelevant Alternatives)
  - Ratio of any two alternatives' choice likelihood is independent of other (irrelevant) alternatives

IIA (Red-Bus/Blue Bus)

- IIA assumes constant substitution rates and leads to inflated shares for similar or identical product offerings
- Because early choice modelers were working on transportation problems, it seemed natural to illustrate the issue using red and blue buses
- Adding a new color bus nearly doubles the predicted share for bus (compared to car, train, walk, and bike)
  - Modelers in the 70s and 80s sometimes included cross-effect terms to slay the IIA dragon

Randomized First Choice (RFC)

- In 1998, most CBC researchers were still using either aggregate logit or latent class MNL
    - Red-Bus/Blue-Bus problem was a big issue with aggregate models
- Bryan Orme proposed RFC in 1998 to penalize products that shared multiple attributes in common

How RFC Works

- The unit of analysis (population, segments, or respondents) is sent on multiple “shopping trips” (iterations), where the unit of analysis makes a “first choice” vote in each trip

- For each shopping trip, the utilities are perturbed by random error (normal, zero-centered)

RFC Example (2 attributes)

Attribute Levels:	Utilities	Iter 1	Iter 2, etc.
Chocolate	0.25	0.25 + -0.21 = 0.04	0.25 + -0.22 = 0.03
Vanilla	-0.25	-0.25 + -0.08 = -0.33	-0.25 + -0.43 = -0.68
Cone	0.30	0.30 + 0.43 = 0.73	0.30 + -0.15 = 0.15
Cup	-0.30	-0.30 + -0.10 = -0.40	-0.30 + 0.48 = 0.18
Bowl	0.00	0.00 + -0.23 = -0.23	0.00 + 0.19 = 0.19
Simulations:			
Chocolate Cone:	0.55	0.77	0.18
Chocolate Cup:	-0.05	-0.36	0.21
Vanilla Bowl:	-0.25	-0.56	-0.49
First Choice:	Choc Cone	Choc Cone	Choc Cup

Green values are normally distributed “attribute-type” error that we draw for each iteration, leading to potentially different first choices across iterations

Correlation Matrix

Correlation in total utility of these 3 product alternatives (across 100,000 simulated iterations):

	ChocCone	*ChocCup*	*VanBowl*
ChocCone	1.000		
ChocCup	0.502	1.000	
VanBowl	0.001	-0.003	1.000

Larger positive correlation between pairs of product alternatives causes them to cannibalize each other more relative to other offerings

How to Decide Variance of the Attribute Error?

- Lower variance leads to “steeper” share results
  - Variance=0 replicates the “first choice” rule!
- Higher variance leads to “flatter” share results
- Sawtooth Software automatically tunes the variance so shares have similar scaling as expected from the share of preference simulation model
  - But you can override with your own tuning

Product-Type Error (Optional)

- Default in RFC is to add only attribute error, causing two identical products to exactly split their shares (relative to having one product)

- An advanced option adds tunable product-type error (to the product utility sum, distributed Gumbel), allowing you to tune the degree of IIA you want in the simulation results
  - Some degree of IIA may, for certain markets, better reflect real-world purchases

- Given that you can tune the contribution of attribute-type and product-type error in RFC, Share of Preference and First Choice rules are special cases of RFC!

HB Draws

- Every iteration in HB produces a “draw” (a proposal) for each respondent’s utilities

- Traditionally, we have averaged those draws (after convergence) to create “point estimates” for each respondent
  - So each respondent has just one set of “summary” utilities

Simulating on HB Draws

- We could use, say, 100 or 1000 draws for each respondent and simulate each respondent's choice 100 or 1000 times!
  - We could apply either the share of preference or first choice simulation rules within each draw

RFC Is Poor Man's Draws

- RFC draws are i.i.d, with no correlation of errors across different levels (within unit of analysis)
  - Two levels of brand (Diet Coke, Coke Zero) get independent attribute-error draws in RFC
- In contrast, individual-level HB draws are not i.i.d and within respondent echo the patterns of correlation from the upper-level covariance matrix
  - So, two SKUs (Diet Coke, Coke Zero) that are different levels of Brand may be positively correlated across draws within the same respondent

RFC Has a Simple View on Similarity

- With RFC, products are only viewed as similar if they share some attribute levels in common
  - Two levels of the same attribute that respondents might perceive as substitutable will still be treated as unique by RFC

Academics (Bayesians) Prefer Draws

- Simulating using the draws is more true to the HB model
  - Averaging across draws (point estimates) is convenient, but not true to HB and its likelihood equation

- Simulating using the draws preserves the proper uncertainty in the data and reflects correlations (within each respondent's draws) as seen in the upper-level covariances

But, Draws Are Less Convenient!

- For practitioners, use of lots of draws can lead to longer simulation runs and larger file sizes
  - RFC is faster, as built into Sawtooth Software's simulation platforms
  - Perhaps 30 or so (evenly sampled) HB draws per person could be used, leading to reasonable file sizes and simulation performance

- Yet, clients like to see one set of utilities for each respondent, not multiple!

So, Which Approach Works Better?

- Practitioners mainly care about the predictive validity of the results, not the elegance of the theory

2016 CBC Modeling Competition

- N=600 calibration respondents (we estimated utilities on 6 attributes for these) for Vacation choices

- N=600 out-of-sample holdout respondents (21 "fixed" holdout CBC questions, each with 4 alternatives)
  - The holdouts included pairs of concepts that were very similar to one another

- The best validation data set we know of for comparing different CBC models and how they deal with IIA!

Predictive Results (Very Little Difference)

	R-Squared Fit to Holdouts
Share of Preference (logit rule) on the Draws	0.9092
RFC on Point Estimates	0.9073
Share of Preference (logit rule) on the Point Estimates	0.9004

We tuned the exponent (scale factor) in all cases to control for differences in scale factor. First choice on the draws is not tunable, so it cannot beat a tuned share of preference on the draws. (Share of Preference with huge Exponent -> First Choice)

We “Tricked” Our Simulator

- We ran simulations on the draws by reading draws into the Sawtooth Software simulator as if each draw was a unique respondent
  - Shares of preference were correct, but standard errors and 95% confidence intervals were incorrect due to inflated sample

Additional Comparisons

- Recently, we revisited this N=600 CBC dataset with 6 attributes, to compare RFC and simulations on the HB draws in terms of:
  - Share Inflation for identical products issue
  - Price demand curves (via sensitivity analysis)

Share Inflation for Duplicate Products

- Scenario #1: simulate 6 product alternatives, where Product 1 gets 3.32% share of preference across N=600 respondents

- Scenario #2: same as above, but with a 7th product that is a duplicate of Product 1

- How will net share for Products 1 & 7 compare to share of Product 1 from the first scenario?

Share Inflation Results

0%	RFC (correlated error applied to all attributes)
0%	HB Draws (first choice rule)
25%	RFC (correlated error turned off for Price)
43%	HB Draws (share of preference rule)
55%	HB Point Estimates (share of preference rule)
94%	Aggregate Logit (share of preference rule)

Additional Notes about Red-Bus/Blue-Bus

- Simulating on individual-level utilities reduces the share inflation problem a lot, but it still exists
  - However, some share inflation probably occurs in real world

- The more products in the simulator and the smaller the shares of preference for each product, the greater the share inflation problem

Price Sensitivity Curve “Kinks”

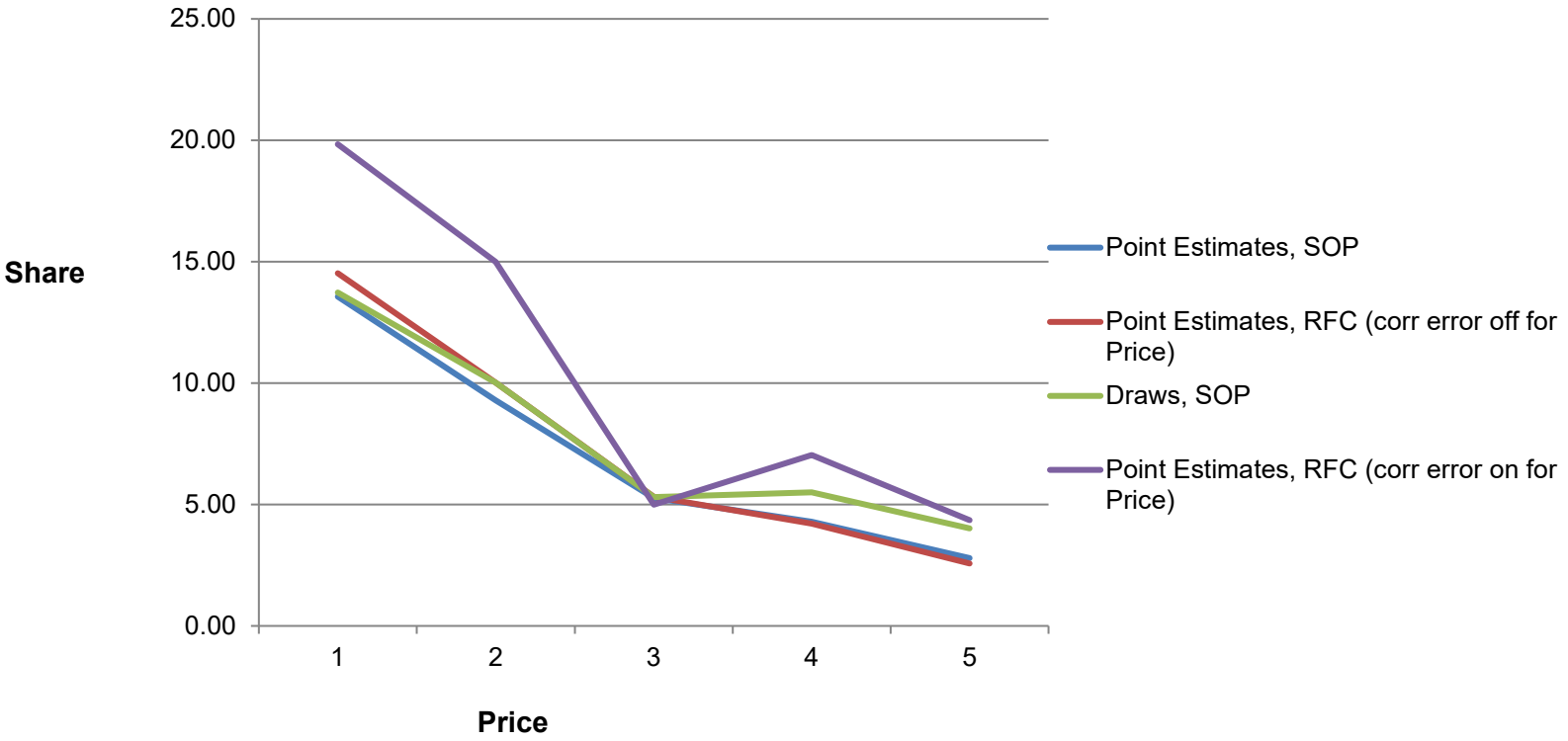
- We often use simulators for developing price sensitivity curves

- We typically set all products at the same (middle) base case price, and this can expose some concerning “kinks” and distortions under different simulation approaches!

Price Sensitivity Curve Setup

- 18 products in the simulation scenario
- All initially set at the middle price point (price 3 out of 5 prices)
- Change the price for the test product, holding all other 17 products at the middle price
  - Record the test product's share at each price point; plot the results

Price Sensitivity Results



What about First Choice on the Draws?

- The distortion “kink” in the price demand curve is even greater than share of preference on the draws
  - A reversal appears, with price 4 preferred to price 3

Another Concern: “Tail Effect” Bias

- HB Draws simulators: be careful about linear coefficient estimation, where you fit a single coefficient to an attribute like Price.
  - A significant % of draws can show reversals on the Price coefficient (positive slope)
  - Strange upward bias in simulating expensive products at highest price points

- “Tail Effect” bias (Karty 2010) greatly reduced when simulating on point estimates, but can still happen!

Key Points to Remember:

- Building simulators on individual-level utilities decreases the IIA problems significantly
  - But, they are still there
  - IIA problems grow as the number of products in the simulation scenario grows
- RFC can further reduce IIA problems
  - And, can be applied to utilities estimated in any way (OLS, aggregate logit, HB-MNL, self-explicated, etc.)

Key Points to Remember:

- Simulating on the HB draws is more defensible and true to HB, but…
  - It takes longer than RFC, with big file sizes (e.g. 1000 respondents with 100 draws is now 100,000 respondents)
  - For the 2016 “prize” CBC data set with six attributes, the predictive results were essentially the same as RFC
  - Some strange “kinks” if running standard sensitivity simulations where all products are initially set at the same price
  - “Tail effect” bias if using linear coefficients for Price, unconstrained
- Note: First choice on the draws is not tunable for scale, whereas Share of Preference on the draws can be tuned (even to match first choice results)

What about Standard Share of Preference?

- If you're building your own market simulators using the Share of Preference (Logit) rule, then you're doing OK…
  - As long as you are using individual-level utilities
  - If you believe that a modest degree of IIA and share inflation behavior is appropriate for predicting the market
  - But watch out for the "IIA-meltdown" that often occurs with big simulation scenarios involving dozens of product alternatives

When NOT to Use RFC

- Two-attribute study with SKU and price
  - RFC thinks that each different SKU is unique (doesn't share a common level)
  - RFC should turn off correction for similarity for Price

- Therefore, no correction for product similarity is being implemented in RFC for SKU & Price!
  - You're getting similar results as Share of Preference, but waiting longer for the RFC iterations!

When to Simulate on HB Draws

- Presenting in front of an academic/Bayesian audience

- 2-attribute studies involving many SKUs and Price might benefit…
  - But we recently did an investigation of 39-SKU and 51-SKU brand/price CBC datasets, and found simulating on the point estimates or the draws gave essentially the same results
  - Nearly same patterns of sourcing (substitution rates among the SKUs) either way
- (But be careful about linear price specifications, for a “tail effect” bias if there are a substantial number of reversed price coefficients!)

Other Advanced Tricks…

…to control for IIA:

- Top N Share of Preference rule (van der Wagt 2014, available in Sawtooth Software's Lighthouse Studio simulator)
- Post Hoc nested share of preference (van der Wagt 2014, also described in Chrzan & Orme's "Becoming an Expert in Conjoint Analysis")
- Nested logit (Lattery 2013)

Thank-You for Listening!