Some thoughts on “Big Data and Marketing Analytics in Gaming: Combining Empirical Models and Field Experimentation” by Nair et al. (2013)

Puneet Manchanda, University of Michigan
Kilts Center Big Data Conference, Booth School of Business, Chicago, Oct 31, 2014
So what’s not to like about this paper?

- Looks at a real problem in an important industry
  - Variants of the problem exist in multiple industries
- Uses many (recent) developments in marketing science to deal with resource allocation in targeted settings
- Uses model based findings to propose a different set of x’s for the setting and
  - Validates them in the field
    - Hallmark of Misra and Nair!
- Has support of a corporate partner
  - Happens less often than it should
More what’s not to like about this paper..

- Straddles the world of academia and of practice

- Showcases marketing science in the world of Big Data
  - Currently (in my opinion) marketing science is very under-represented

- Casino industry is highly promotion sensitive
  - So very impressive to find 6.7% increase (R$4.57/R$68.07) or $1mm - $5 mm incremental return

- Moral of the story – it’s all in the data generating process
The generic sales response model..

\[ y_{it} = f \left( x_{it} \mid \beta_i \right) \]

Unit of aggregation \( i \)
(account, store, territory, customer)

Response Parameters

Inference focuses on conditional model: \( y \mid x \)
And the standard solution ...

The assumption here is that marginal distribution of $x$ ($x_{it} | \theta$) provides no information about the response parameters.

So the likelihood factors as follows

$$
\ell(\{\beta_i\}, \theta) = \prod_{i,t} p(y_{it} | x_{it}, \beta_i)p(x_{it} | \theta)
$$

$$
= \prod_{i,t} p(y_{it} | x_{it}, \beta_i) \prod_{i,t} p(x_{it} | \theta).
$$
But in data-rich settings

- x values often set with (partial) knowledge of response parameters
- So model needs to be modified as (Manchanda, Rossi, Chintagunta JMR 2004)

\[ y_{it} | x_{it}, \beta_i, \text{ and } x_{it} | \beta_i, \tau \]

- This allows us to obtain unbiased parameter estimates
  - In addition, the use of information in x about parameters can “sharpen” the estimates
A different solution here

- Two institutional features (IF) of the data used to address issue
  - Value of corporate partner

- IF 1: The data generating process of $x$ is known (almost perfectly)

- $x$ are a function of past behavior ($z$) and demographics ($d$)
  - More important, it turns out that $x$ are not a function of response parameters
A different solution here (contd.)

- So, given \( z \) and \( d \), each observation (consumer-month) can be assigned to segment \( s \) in a deterministic manner
  - Assignment does not consider any unobservables so unlike scoring function approach
  - Analysis is conditional on consumer belonging to segment \( s \) at time \( t \)
  - Allows for within-consumer (time-varying) heterogeneity

- IF 2: Assignment of \( x \) within segment \( s \) is randomly provided to a subset of consumers in \( s \)
  - In essence, the response to \( x \) is estimated in a series of iid draws from within segment \( s \)

- Thus estimates of the response parameters for a given segment are unbiased
Assumptions and boundary conditions

- Assignment to segment $s$ is based (partly) on $z$ (past behavior)
  - But past behavior can be a function of responsiveness to promotions (as they affect the propensity to visit, play, spend etc.)
  - So is segment membership completely uncorrelated with response parameters?
    - If not, then response parameters could be biased even within segment e.g., for heavy play segments, promotions are always high (p. 9), leading to spurious correlation between volume of play and promotion
- Random assignment within segment to conditions of no promotion versus promotion will “unconfound” this
  - Will help if the authors can show these patterns in the data
Assumptions and boundary conditions

- Response parameters for segment $s$ are invariant to who is in segment $s$
  - In other words, if my $(z, d)$ change and I move from $s_1$ to $s_2$, then I automatically get assigned $s_2$’s response parameters
    - Can we get a sense of the movement of individuals across segments?
  - The casino industry actually tries to move you to more active (valuable) segments the more it knows about you
    - So while hope is to change responsiveness, that may or may not happen

- The proportion of consumers assigned to a promotion within a segment needs to be “small”
  - If not, then repetitions are not iid and
  - Effects such as learning etc. can kick in, leading to non-stationary response parameters (within segment)
  - Great if authors could share more data on these proportions
Assumptions and boundary conditions

What about strategic behavior?

As the authors note, customers form expectations vis-à-vis promotions/rewards

- Implication is that promotions need to reach some threshold before response is seen i.e., response curve may be highly non-linear
- Does the casino company already adjust for that (while the model doesn’t)?
- Probably not an issue in the field experiment as it stays within range of data (and temporal duration of data is short)

How important is the role of state-dependence?

- Could manifest itself in satiation, addiction etc. (Narayanan & Manchanda 2012 QME), leading to changing promotion response over time
- Current approach “force-fits” this individual level evolution by moving him/her to “appropriate” segments over time
Assumptions and boundary conditions

- How much do other context effects matter?
  - Within month variation (weekend, payday etc.), seasonality (Field Experiment in Q3)
  - Competitive promotions
  - ....
- But at this scale, average effect over segment-month (as reported here) is a good starting point
Minor quibbles and questions

- **Nested Logit structure**
  - Does it map to consumer decision making process (even though it’s an “as-if” model)?
  - Are promotions seen as discrete choices or as dollar values?
  - Can a consumer really choose from multiple promotions for a given property (and multiple properties) for a given month?
    - Not possible in the field experiment (p. 31)

- **Paper notes (p.7, p. 9) that current promotions are based on RFM**
  - Is that only across segments or within as well?
  - Great if the authors could show the raw data patterns
Minor quibbles and questions

Figure 4: Nesting Structure Used in Model Setup

Level 2: Visit and Brand Choice Decision – Customer decides to visit a specific MGM Resorts brand or not visit any property

Customer Decision process

Level 1: Promotion Decision - Customer decides which promotion to accept or alternately to visit without accepting the promotion

- Visit Bellagio
  - Accept promo 1
  - Accept promo 2
  - Accept promo 3
  - Accept promo 4
- Visit MGM Grand
  - Visit Bellagio w/o accepting promotion
  - Accept promo 4
- Visit Aria
  - Accept promo 5
- Visit Brand...
  - Accept promo 6
- Do Not Visit MGM Property
Minor quibbles and questions

- Is this really an application of Big Data?
  - Is no. of segments, coefficients etc. what decides Big versus Small data?
    - CPG firms run very large scale models at SKU level
    - Pharma companies run large non-linear models for 1mm+ physicians
    - Targeting here is quite macro
      - Segments in order of 100s – consumers in order of 1,000,000s
  - Caveat: Big Data is like teenage sex
  - Opportunity for authors to take a stand on definition

- How representative is the casino industry?
  - 15-20 year history of very detailed data collection and analytics
  - Random assignment within segments is unusual in most other settings
  - Highly promotion sensitive customer base
  - Is there much more upside with respect to promotion?
The bigger picture

- Authors conclude the paper with some valuable tips on how to get analytics to work inside the organization
  - Adding to that, in my experience, top management involvement is critical
  - Would also have been nice to get some detail on the cost of data collection & cleaning (authors note that is a very painstaking process), running experiments, data analysis, optimization etc.
    - If the time it takes to do this on a regular basis > decision-making cycle, then need some shortcuts

- The role of structure
  - If objective is prediction (and profit), how much worse off are we running (model free) large scale random experiments (e.g., A/B testing in each segment)?
    - This is especially relevant for digital businesses as cost of experimentation is low

- How do we foster an environment where more academic researchers can engage with companies at this level of rigor and relevance?