

Clever Randomization and Ensembling Strategies for Accommodating Multiple Data Pathologies in Conjoint Studies

Marc Dotson (BYU) and Jeff Dotson (BYU)

Marketing Science

June 11, 2020

Conjoint Analysis in Marketing

- ▶ Products are defined by **attributes**, each with a number of **levels**
- ▶ Individuals choose from among a number of **alternatives**
- ▶ We infer **respondent-level preferences** for each attribute level
- ▶ We make use of preference estimates to **predict demand** for potential products compared with the competition
- ▶ These **market simulators** can inform new product development, pricing, product line optimization, and go-no-go decision-making

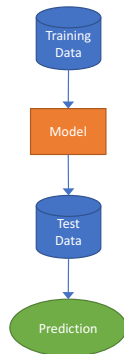
Current Approach to Conjoint Prediction

1. Training dataset

2. Estimate HBMNL

3. Simulate outcomes
using new data

4. Generate predictions



Data Pathologies

Pathology: Structural and functional **deviations from the normal** that constitute disease or characterize a particular disease

Data Pathology: Any **respondent-level behavior** that deviates from the **normative model** of consumer choice, thus impeding accurate out-of-sample prediction

Normative Model of Consumer Choice

For a given individual and choice task:

$$y \in \{1, 2, \dots, J\}$$

$$U_j = V_j + \varepsilon_j$$

$$V_j = \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_k x_{kj}$$

$$\varepsilon_j \sim \text{iid Gumbel}$$

Across respondents:

$$\beta \sim N(\bar{\beta}, \Sigma)$$

Models of Data Pathologies

- ▶ Many innovations in conjoint focus on addressing **specific data pathologies**
- ▶ These models fit better and provide marginal improvements in predictive fit
- ▶ Although prospectively useful, these models are **rarely used in practice**

Examples of Data Pathologies

- ▶ Attribute Non-Attendance
- ▶ Screening Rules
- ▶ Poor Respondent Quality
- ▶ Respondent Fatigue
- ▶ Alternative Decision Rules (e.g., Regret Minimization)

Difficulties in Practice

- ▶ They are theoretically and computationally complex (i.e., difficult to understand and time-consuming to estimate and simulate)
- ▶ We don't have high-quality commercial software that can be used to fit these models
- ▶ Each model deals with a single data pathology
 - A priori how do I know which pathology will prove problematic (i.e., which model should I fit)?
 - What if multiple pathologies are present in a single dataset?

Ensemble Approaches to Prediction

- ▶ With a few exceptions, most conjoint projects are still conducted using a **single model and dataset** for both inference and simulation
- ▶ However, **ensemble approaches** currently dominate the world of competitive prediction (e.g., Kaggle, Netflix Prize, etc.)

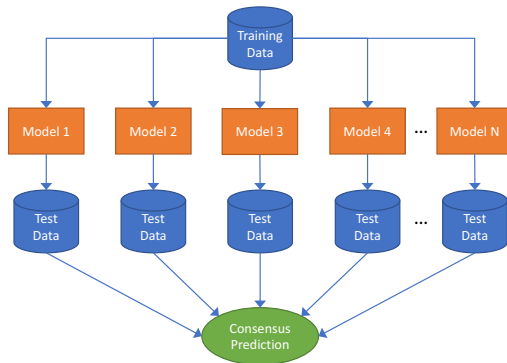
An Ensemble Approach to Prediction

1. Training dataset

2. Estimate different types of models

3. Simulate outcomes on new data

4. Generate a consensus prediction



Why Do Ensembles Perform So Well?

- ▶ They strike an optimal trade-off on the bias/variance frontier
 - Wisdom of the Crowd
- ▶ They allow for efficient exploration of the model space
 - All models are wrong, but **some** models are useful
- ▶ They provide a hedge against **data conditions that are pathological** with respect to a single model (our hypothesis)

Our Proposed Approach

- ▶ Create an ensemble where **diversity** is created through various forms of **(clever) randomization**
- ▶ Induce diversity in the ensemble by thinking about how **specific forms of randomization** are related to **specific data pathologies**
- ▶ Generate a **consensus prediction** from the ensemble

Note: Our focus in this study is to improve **out-of-sample predictive validity!**

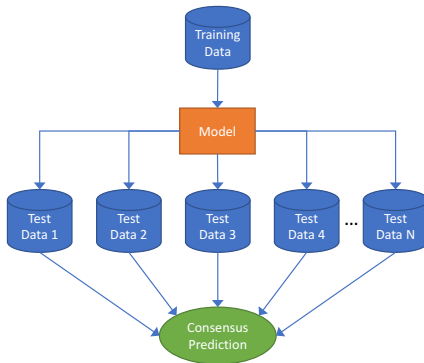
An Ensemble Approach to Conjoint Prediction

1. Training dataset

2. Estimate HBMNL

3. Simulate outcomes using random versions of holdout data

4. Generate a consensus prediction



Pathology 1: Attribute Non-Attendance (ANA)

Attribute Non-Attendance: Respondents ignore subsets of attributes when making decisions (i.e., $\beta = 0$ for all levels of the attribute)

Randomization Strategy for the Ensemble: Create an ensemble where we randomly set the coefficients for a full attribute to 0

Ensemble Algorithm for Attribute Non-Attendance

1. Estimate an HB MNL on training data
2. Loop over respondent-level estimates ($\hat{\beta}_h$)
 - randomly select an attribute
 - with probability θ set all coefficients for that attribute to 0
3. Predict first choices (e.g., max utility) for each choice set in the out-of-sample holdout (test) data
4. Repeat 2 and 3 many times
5. Generate a consensus (e.g., most commonly selected) prediction

Pathology 2: Screening Behavior

Screening Behavior: Respondents use certain attribute levels to screen out alternatives from consideration (i.e., $\beta \approx -\infty$ for levels being screened on)

Randomization Strategy for the Ensemble: Create an ensemble where we randomly set the coefficients for attribute levels to $\approx -\infty$

Ensemble Algorithm for Screening Behavior

1. Estimate HB MNL on training data
2. Loop over respondent-level estimates ($\hat{\beta}_h$)
 - randomly select an attribute level
 - with probability ψ set the coefficient for that attribute level to $\approx -\infty$
3. Predict first choices (e.g., max utility) for each choice set in the validation data
4. Repeat 2 and 3 many times
5. Generate a consensus (e.g., most commonly selected) prediction

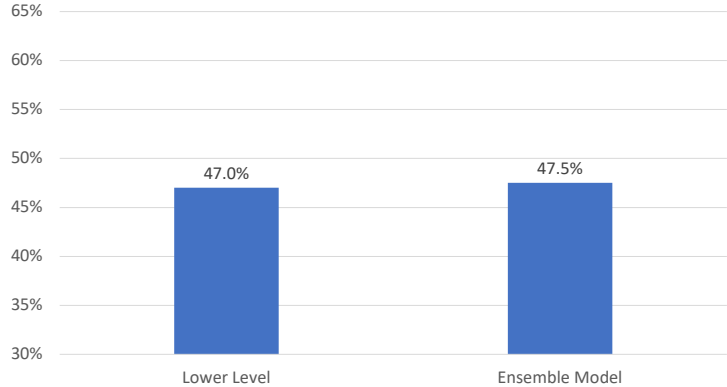
Joint Ensemble Algorithm for Attribute Non-Attendance and Screening

1. Estimate HB MNL on training data
2. Loop over respondent-level estimates ($\hat{\beta}_h$)
 - randomly select an attribute level
 - with probability ψ set the coefficient for that attribute level to $\approx -\infty$
 - randomly select an attribute
 - with probability θ set all coefficients for that attribute to 0
3. Predict first choices (e.g., max utility) for each choice set in the validation data
4. Repeat 2 and 3 many times
5. Generate a consensus (e.g., most commonly selected) prediction

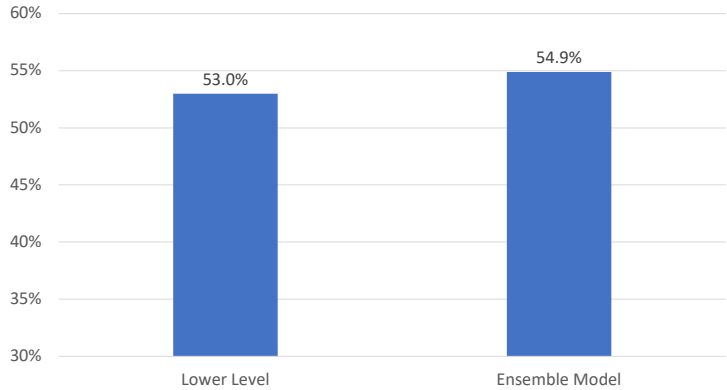
Simulation Study

- ▶ Simulate data with the presence or absence of pathologies: attribute non-attendance and/or screening
- ▶ 500 respondents; 12 choice tasks; 6 attributes
- ▶ Generate prediction from standard HBMNL results (lower level) and the ANA, Screening, or joint ANA + Screening Ensemble

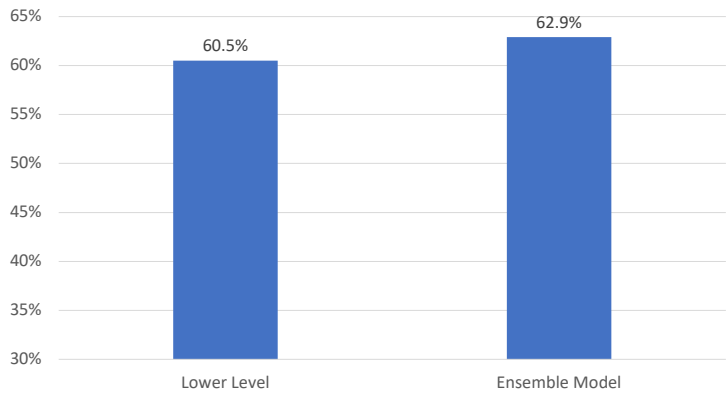
Simulated Data: No ANA + No Screening



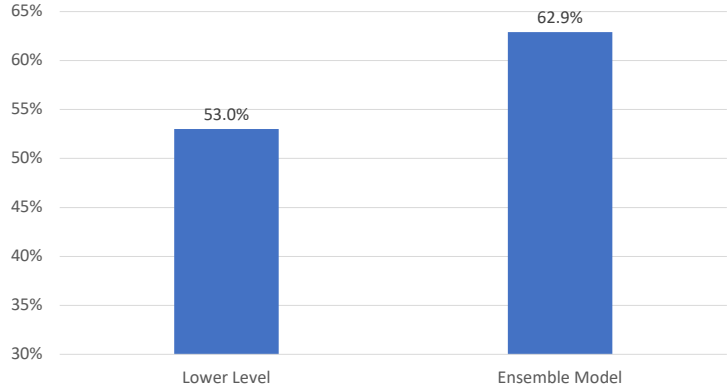
Simulated Data: ANA without Screening



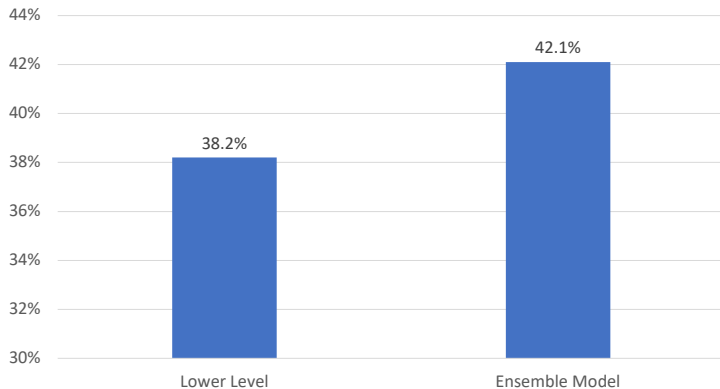
Simulated Data: Screening without ANA



Simulated Data: ANA + Screening



Performance on Interior Paint Data



Another Ensemble Approach to Conjoint Prediction

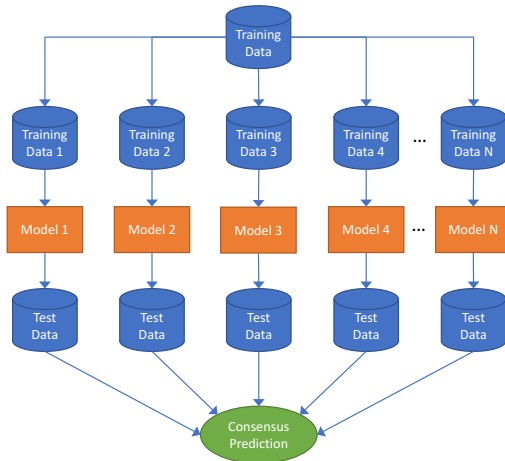
1. Training dataset

2. Create random subsets of the data

3. Estimate HB MNL

4. Simulate outcomes using random versions of holdout data

5. Generate a consensus prediction



Next Steps

- ▶ Utilize computational advantages of Hamiltonian Monte Carlo via Stan
- ▶ Explore clever randomization strategies on test data
- ▶ Include more pathologies
- ▶ Better aggregation of prediction (e.g., model stacking via a meta-learner rather than consensus predictions)

Conclusion

- ▶ Cleverly designed ensembles of models appear to help hedge against data pathologies and improve prediction
- ▶ The ensemble approach is especially powerful in the presence of multiple data pathologies
- ▶ Ensembles for inference: Can we learn about the severity of a particular pathology by examining the relative fit of ensemble members as part of a meta-learner?

Thank You!

Questions? Contact me at marc.dotson@byu.edu.