**Accommodating Multiple Data Pathologies in**

**Conjoint Studies via Clever Randomization and Ensembling**

***Jeffrey P. Dotson***

*Brigham Young University*

***Roger A. Bailey***

*The Ohio State University*

***Marc R. Dotson***

*Brigham Young University*

**1 Introduction**

Ensemble-based approaches currently dominate the world of competitive out-of-sample prediction. From Kaggle to the Netflix Prize, the predictive power inherent in using many models overshadows prediction reliant on the performance of a single model. The primary reason ensembles predict so well is that they serve as a hedge against model misspecification. Since we have uncertainty about the correct model for any given context, running many models and producing a consensus is a simple yet powerful way to improve predictions.

In the world of conjoint, most studies are conducted using a single model. When the aim of a conjoint study is solely inference and not prediction, a single-model approach is arguably best. The academic literature for conjoint is filled with models designed to improve inference, especially when respondents behave in ways that are “pathological” to the standard model. However, there are three reasons to argue for an ensemble-based approach to conjoint analysis. First, the end goal of many conjoint studies is prediction in the form of accurate market simulations. Second, we still have uncertainty about the correct model for any given conjoint study. Third, there is no single model that accounts for all the respondent behaviors that result in the “data pathologies” that have been addressed separately in the literature.

The remainder of the paper will be organized as follows. In Section 2, we walk through ensemble approaches to prediction. In Section 3, we detail our ensemble approach to conjoint analysis. In Section 4, we provide results from simulation studies and an empirical application. In Section 5, we conclude.

**2 Ensemble Approaches to Prediction**

Before walking through ensemble approaches to prediction, it’s helpful to review the terminology commonly used in this space. The single-model approach to prediction is illustrated in Figure 1. The steps are to, first, specify the data used to train (i.e., estimate) the model; second, train the model; third, simulate outcomes using parameter estimates and test data (e.g., hold-out tasks or hold-out respondents); and fourth, use these simulated outcomes along with the test data to compute a prediction (e.g., hit rates).

**Figure 1**

A single-model approach to prediction.

**

The single-model approach to prediction is the standard for conjoint studies. One notable exception is Kevin Lattery’s Sawtooth 2015 presentation and paper, “A Machine Learning Approach to Conjoint.” In that paper, Lattery implemented an ensemble approach to prediction for conjoint analysis, which is illustrated in Figure 2. Here we can see that many models are fit and many predictions are calculated, each model using its own randomly selected test data. Finally, a consensus prediction is formed via aggregating the separate predictions (e.g., averaging predictions or taking the modal prediction).

**Figure 2**

An ensemble approach to prediction.



Lattery finds substantial improvements in prediction as the number of models in the ensemble increases. However, the ensemble isn’t theoretically grounded. Furthermore, the approach to prediction is fairly non-standard. A more general ensemble approach to prediction is illustrated in Figure 3. The additional step is to create a random subset of the data to serve as the training data for each of the models in the ensemble. Having separate training and test data for each of the models in the ensemble is the standard approach to ensembles.

**Figure 3**

An alternative ensemble approach to prediction.



Even with a non-standard ensemble approach to prediction and no theoretical justification for the ensemble, Lattery still found improvement in out-of-sample prediction. As stated previously, ensemble approaches to prediction are powerful because they serve as a hedge against model misspecification. This is often justified by ensembles striking an optimal balance on the bias/variance frontier. For example, in Figure 4 we can see two hypothetical targets. On the left we have low variance, high bias performance that represents using a single albeit mis-specified model for prediction. On the right we have a high variance, low bias performance that represents the ensemble approach to prediction. The ideal is to have low bias and minimal variance. Our aim is to approach this ideal by using an ensemble that is theoretically justified.

**Figure 4**

Bias variance trade-off in ensembles.



**3 Accommodating Data Pathologies**

Our ensemble-based approach to prediction, like Lattery’s, is non-standard. However, our approach differs in that we build an ensemble that is theoretically justified. In particular, we introduce randomization in our ensemble that is “clever” insofar as the randomization accounts for a potential data pathology. As noted above, much work has been done to build models that separately account for respondents producing data that is pathological to the standard model (i.e., can’t be accounted for by the standard model and thus impedes prediction). Our use of an ensemble and clever randomization allows us to accommodate multiple data pathologies.

The standard model is a lower-level random utility model with an upper-level model over preferences. Consumers are able to assess the “utility” of each alternative in a choice set and pick the alternative that provides the greatest level of utility. Utility itself is made up of two components: A deterministic component and a random (to the researcher) component, where the deterministic component is expressed as a (linear, compensatory) function of the design of the alternative and the random component is assumed to come from an independent and identically distributed Gumbel distribution. Finally, while we estimate preferences at the individual level, we assume that the preferences of all individuals are drawn from a common multivariate normal distribution.

Development in the academic conjoint literature has focused on addressing specific data pathologies separately (e.g., attribute non-attendance, screening rules, poor respondent quality, non-IIA choice behavior, respondent fatigue, and alternative decision rules). These models fit the data better and provide marginal improvements in prediction. Although prospectively useful, especially in terms of inference, these models are rarely used in practice for three reasons. First, they are theoretically and computationally complex (i.e., difficult to understand and time-consuming to estimate and simulate). Second, we don’t have high-quality commercial software that can be used to fit these models. Third, each model deals with a single data pathology. The challenge of model misspecification persists wherein, a priori, it is hard to know which pathology will prove problematic (i.e., which model should be fit). This problem is further complicated if, as might be expected, multiple pathologies are present in a single dataset.

In our ensemble-based approach to prediction, we don’t have to fit complicated models, but we do need to compute a lot of predictions! The trade-off is between model complexity and computational intensity. Figure 5 illustrates our approach. Note that this is non-standard, in that we have both a single training dataset and a single model. The randomization we introduce is at the level of the randomly selected test data. We will accommodate two data pathologies, attribute non-attendance and screening behavior, separately and jointly.

**Figure 5**

Our ensemble approach.



**Pathology 1: Attribute Non-Attendance**

Attribute non-attendance is when respondents ignore subsets of attributes when making decisions (i.e., part-worths are 0 for all levels of the attribute). To accommodate this data pathology, we randomly set the part-worths for all levels of an attribute to 0 across test datasets. To be clear, we implement the following:

1. Estimate an HB MNL on training data
2. Loop over respondent-level part-worth estimates
   * randomly select an attribute
   * with a given probability, set all part-worth estimates for that attribute to 0
3. Predict first choices (e.g., max utility) for each choice set in the 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 is when respondents use certain attribute levels to screen out alternatives from consideration (i.e., part-worths are approximately negative infinity for all levels being screened on). To accommodate this data pathology, we randomly set the part-worths for levels to approximately negative infinity across test datasets. To be clear, we implement the following:

1. Estimate an HB MNL on training data
2. Loop over respondent-level part-worth estimates
   * randomly select an attribute level
   * with a given probability, set the part-worth estimate for that level to approximately negative infinity
3. Predict first choices (e.g., max utility) for each choice set in the test data
4. Repeat 2 and 3 many times
5. Generate a consensus (e.g., most commonly selected) prediction

**Joint Ensemble for Pathologies 1 and 2**

To accommodate for both data pathologies, we implement the following:

1. Estimate an HB MNL on training data
2. Loop over respondent-level part-worth estimates
   * randomly select an attribute level
   * with a given probability, set the part-worth estimate for that level to approximately negative infinity
   * randomly select an attribute
   * with a given probability, set all part-worth estimates for that attribute to 0
3. Predict first choices (e.g., max utility) for each choice set in the test data
4. Repeat 2 and 3 many times
5. Generate a consensus (e.g., most commonly selected) prediction

**4 Simulation Studies and Empirical Application**

Four simulation studies demonstrate the potential of our ensemble approach to prediction. Figure 6 shows that when neither of the two data pathologies are present in the simulated data, the lower-level model (i.e., the standard model) and our ensemble approach predict about the same. Figure 7 shows that when attribute non-attendance is present but screening is not, our ensemble approach slightly out predicts the standard model. This is repeated with Figure 8 when screening is present but attribute non-attendance is not. However, Figure 9 clearly demonstrated the benefit of the approach as we see a large jump in predictive ability for our ensemble approach when both of the data pathologies are present in the simulated data.

**Figure 6**

Simulated Data: No Attribute Non-Attendance + No Screening.



**Figure 7**

Simulated Data: Attribute Non-Attendance without Screening.



**Figure 8**

Simulated Data: Screening without Attribute Non-Attendance.



**Figure 9**

Simulated Data: Attribute Non-Attendance + Screening.

****

Figure 10 shows that the improvement in prediction increases as more test datasets following the theoretical justification outlined above are added to the ensemble.

**Figure 10**

Predictive fit as a function of ensemble size.



Finally, Figure 11 shows that the improvement in prediction for our ensemble approach over the standard model is clearly present for real data. Interpolating from our simulation experiments, this is most likely attributed to both of the data pathologies being present in the data, a condition that isn’t accounted for by the more attribute non-attendance and screening models separately.

**Figure 11**

Performance on actual data.

****

**5 Conclusion**

When the goal for a conjoint study is out-of-sample prediction, there is great potential in an ensemble approach. The benefit of our ensemble approach is its construction is theoretically justified, simple to implement, and performs especially well when multiple data pathologies are present in a dataset.

There are a variety of next steps to consider. Our stylized, non-standard approach where clever randomization is induced only for the test data should be expanded to allow for randomization of training data and the training of many models. This will necessitate faster or more efficient computation. Given the benefit we’ve seen with accommodating only two data pathologies, more pathologies need to be considered and accounted for. Finally, introducing smarter ensembles and better prediction aggregation can only improve the approach, especially if the method of aggregation allows us to retain the benefits of inference and not simply produce improved predictions.