Tree-based Marketing Mix Models

Hartaig Singh, MS | Medidata Solutions, a Dassault Systèmes Company
Vikas Hiremath, MS | Medidata Solutions, a Dassault Systèmes Company

MEDIDATA

Key Takeaways

Random forest requires significantly less assumptions/parameters and prior information–drastically reducing turnaround time and modeler discretion.

Random forests are much better at estimating true response curve fits.

Model performance is not related to quality of response curve fit.

Random forest feature importance may be correlated with quality of response curve fit–a leading indicator for response curve accuracy.

Introduction

Marketing mix models are used to assess the impact of multiple promotional channels within a company/brand on ROI.

Bayesian regressions combined with nonlinear transformations and ad stock decay are typical models.

This requires significant modeler discretion, especially on the priors, thus preventing the process from scaling.

Random Forests by design capture non-linearity in response curves, are better at handling a large number of features and collinearity, and identify interplay between media channels without having to specify this explicitly.

This removes the need to manually define forced saturation transformations and minimize the danger of mis-specifying interactions between different channels.

These advantages aid in reducing model complexity and modeler discretion, enabling more granular level insight, scalability, and automation.

Objective

The goal of this analysis is to test the efficiency of random forests over traditional Bayesian regressions.

We test this by observing each model's:

1. Ability to capture saturation via response curves
2. Number of assumptions/judgement calls
3. Number of parameters needed to define
4. Predictive and curve fit performance
5. Flexibility with prior information

Methods

We simulated an HCP specialty, month-level data set with three media channels using specified hill parameters to force c and s shaped curves. The data set also contains three different specialties (segments) each with 120 time points. The dependent variable is the sum of the outputs for each hill equation across each segment.

We fit one random forest model and four different Bayesians regressions. No hyperparameter tuning was performed on the random forest model–default sklearn random forest regressor hyperparameters. For each regression, we vary the saturation parameter's prior assumptions. All models were set up so they capture carryover and saturation effects by segments.

Since we specified the simulation, true segment-level responses and parameters are available. We compare this against the estimated response curves of each model.

Models

Model Component	Random Forest Fit	Bayesian Regression Fits
Main/fixed effects	Media channels (3)	Media channels (3) + Intercept (1)
Dependent variable lag	Lagged dependent variable up to three months (3)	Lagged dependent variable up to three months (3)
Adstock	Lag each media channel for up to three months (9)	Adstock decay for each media channel (3)
One-hot encoding	One-hot encode each specialty (3)	Not needed
Interactions	Included by design	Not defined
Random effects	None	Slopes and intercepts within specialties for each media channel (12)
Non-linearity/Saturation effects	Included by design	Saturation effects for each media channel at the specialty level using hill (18) or logistic (9)
Priors	None	Assume half-normal priors on all coefficients and beta priors on adstock

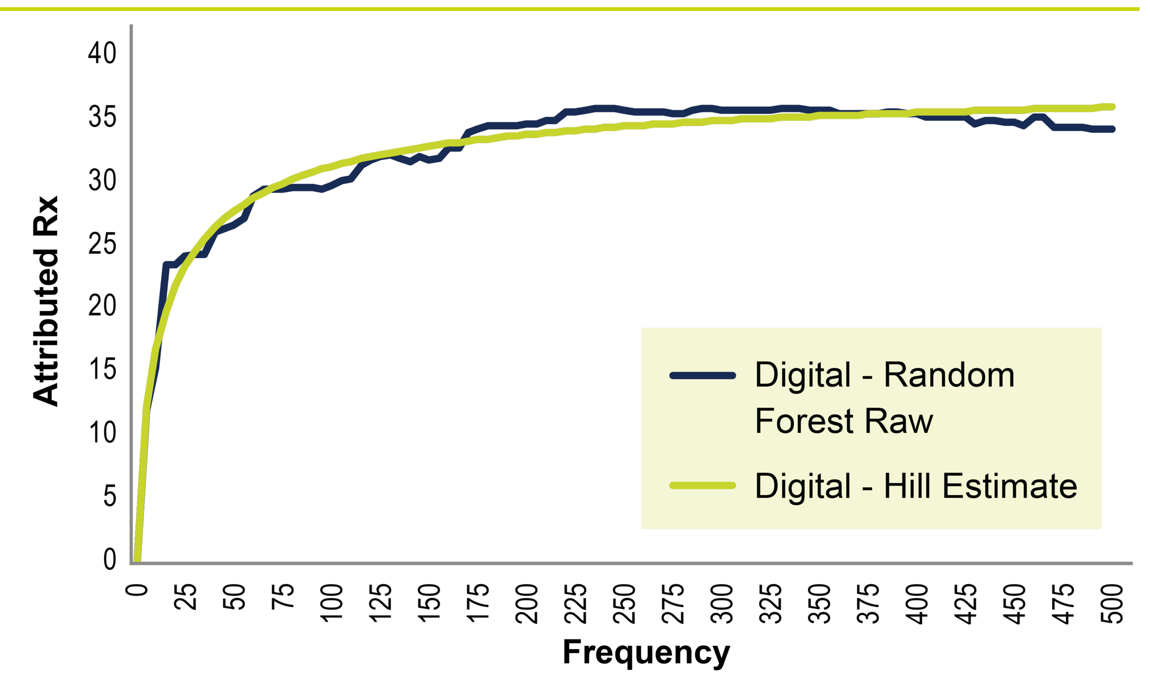
Results

Model	Num. of Guesses	Num. of Params	Model Fit	Curve Fits (Avg. RMSE of actual vs estimate)
Random Forest	4*	18	R2: 98.9% RMSE: 4.4	Email: 0.290 Phone: 0.166 Digital: 0.141
Bayesian Regression: Model 1 Logistic : $\lambda \sim gamma(3,1)$	29**	31	R2: 97.2% RMSE: 0.03	Email: 0.297 Phone: 0.364 Digital: 0.324
Bayesian Regression: Model 2 Hill : $\alpha \sim beta(2,2),\ \gamma \sim gamma(3,1)$	38**	40	R2: 99.5% RMSE: 2.9	Email: 0.295 Phone: 0.364 Digital: 0.324
Bayesian Regression: Model 3 Hill : $\alpha \sim HN(0,10),\ \gamma \sim HN(0,10)$	38**	40	R2: 99.5% RMSE: 2.9	Email: 0.293 Phone: 0.364 Digital: 0.323
Bayesian Regression: Model 4 Hill : $\alpha = 1,\ \gamma \sim N(0,10)$	38**	40	R2: 79.2% RMSE: 19	Email: 0.293 Phone: 0.362 Digital: 0.319

** Lookback length for each channel and DV, ** Lookback length, priors for all media/segment coefficients, saturation, and adstock terms*

Random Forest Response Curve Generation

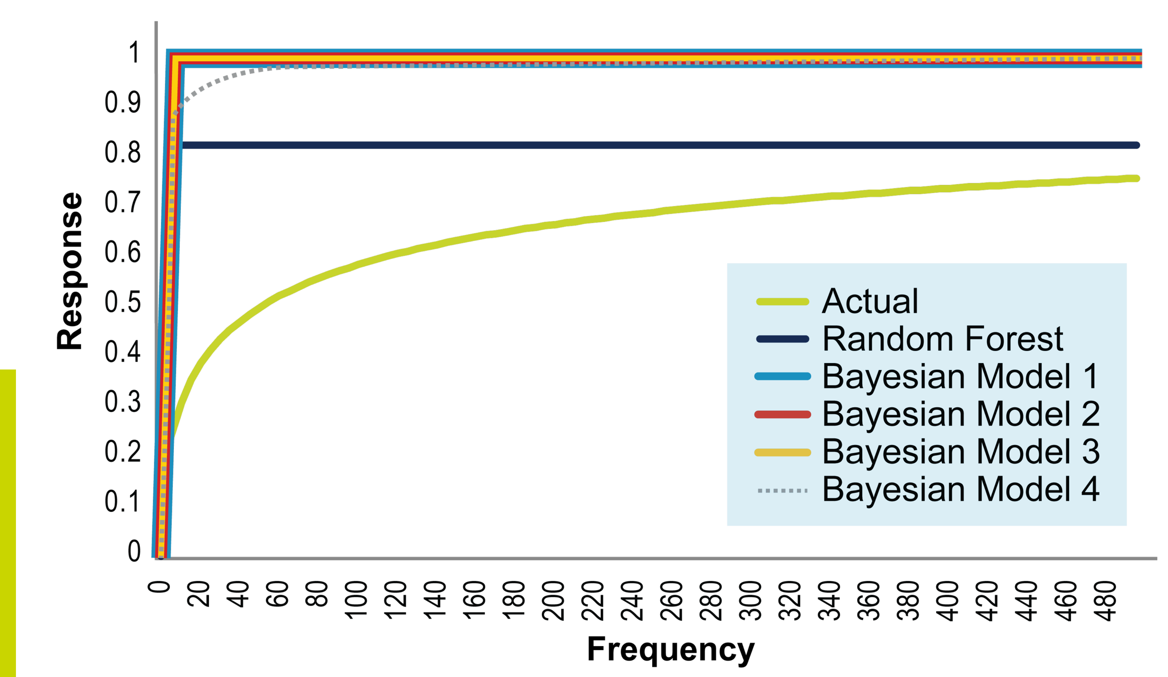
Figure 1.
Blue shows the average predictions for each frequency minus the predictions when digital equals zero. Green is a hill curve fitted over blue.



Digital Response on Neurologists

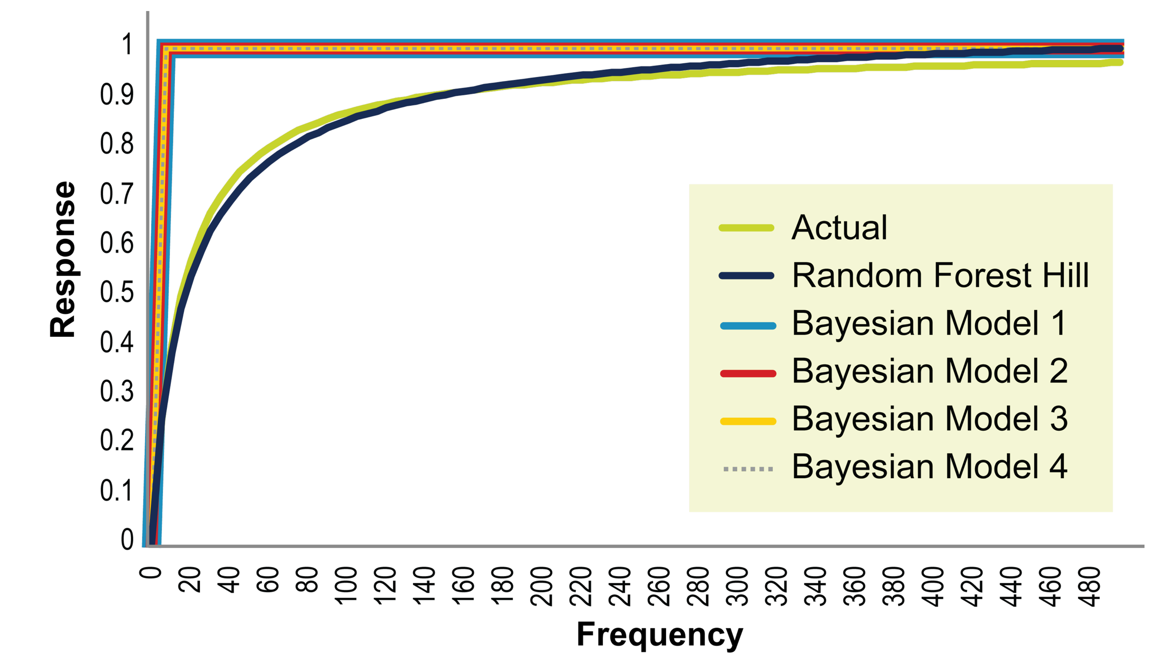
Figure 2.
True responses compared with model estimates. Responses were normalized to standardize scale. The purpose is to see if shape is being estimated well.

Digital impact among neurologists is low (deliberately set this). We hypothesize random forests will not fit responses well for low importance/signal channels–targeting low impact audiences may not be necessary anyway



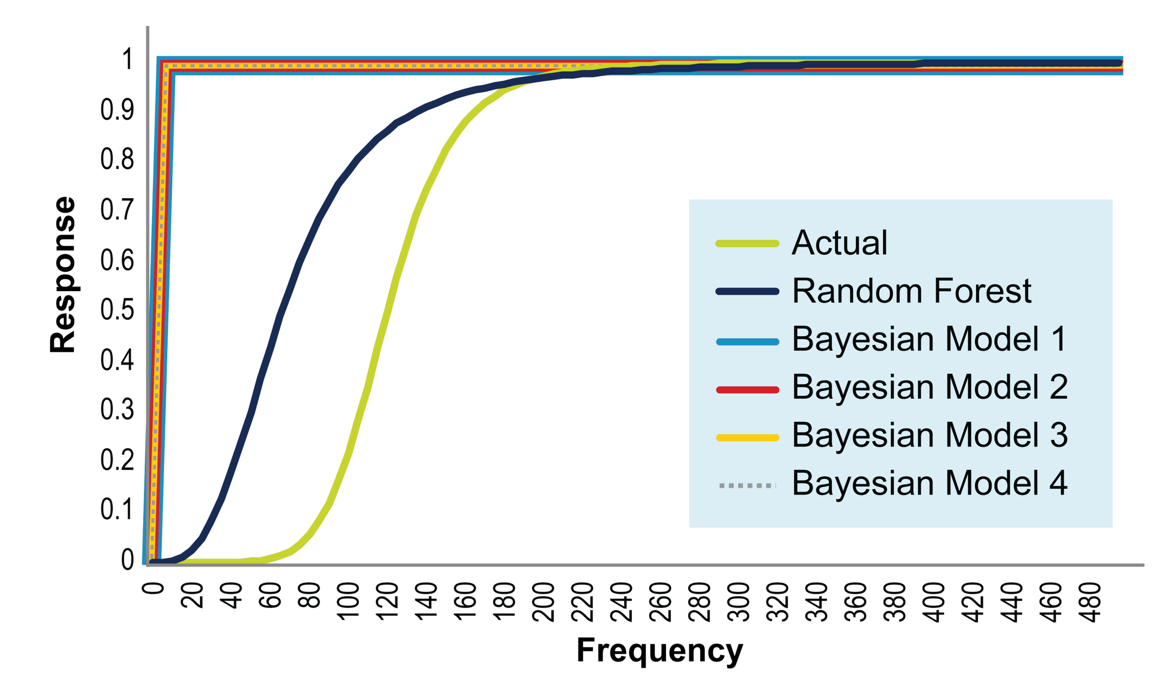
Digital Response on Hematologists

Figure 3.
True responses compared with model estimates. Responses were normalized to standardize scale. The purpose is to see if shape is being estimated well.



Digital Response on Oncologists

Figure 4.
True responses compared with model estimates. Responses were normalized to standardize scale. The purpose is to see if shape is being estimated well.



Conclusion

Varying the prior distribution still estimates similar response curves, showing the need to carefully specify and define informative priors.

Figure 2 shows model performance is not always related to true underlying promotional response.

Figures 3, 4, and table show random forest's ability to capture segment-level responses and true fits significantly better than Bayesian models.

Random forest achieves more reasonable promotion response fits while needing to specify/assume minimal parameters.

A limitation of random forest is their inability to include priors for cases where accurate estimates of responses are available.

In cases where minimal/weak prior information is available, random forest may be a stronger first attempt.

References/Acknowledgements

"Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects" – Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, Jim Koehler, Google Inc. 14th April 2017
Eric Yang, PhD – Medidata Solutions, a Dassault Systèmes Company
Hugo Catch, MS – Medidata Solutions, a Dassault Systèmes Company