

Bayesian Hill Mixture Models for Heterogeneous Consumer Response in Marketing Mix Modeling

January 2026

Abstract

Marketing Mix Models (MMM) are widely used for measuring advertising effectiveness, yet standard implementations assume homogeneous consumer response to marketing spend. This paper proposes a Bayesian mixture of Hill saturation functions to capture heterogeneous response patterns across latent consumer segments. We implement the model using NumPyro with automatic prior scaling and ordered constraints for identifiability. Experiments on simulated data demonstrate that mixture models achieve superior predictive accuracy compared to single-curve baselines, with proper recovery of adstock parameters.

1. Introduction

Marketing Mix Modeling (MMM) enables organizations to quantify the effectiveness of marketing investments and optimize budget allocation. Modern implementations typically employ Hill saturation functions to capture diminishing returns and geometric decay adstock transformations to model carryover effects (Jin et al. 2017; Chan and Perry 2017).

A critical assumption underlying standard MMM implementations is that all consumers respond identically to marketing stimuli. In reality, different segments—heavy versus light buyers, brand loyalists versus switchers—exhibit heterogeneous response patterns (Wedel and Kamakura 2000; Allenby and Rossi 1998). Aggregate response curves represent weighted averages that may mask segment-specific behaviors and lead to suboptimal allocation decisions. Recent work has also shown that nonlinear effects in MMM may be artifacts of model misspecification (Dew et al. 2024), further motivating flexible mixture approaches.

This paper addresses these limitations by proposing a Bayesian mixture of Hill saturation functions. The model simultaneously estimates latent segment membership probabilities, segment-specific saturation parameters, and shared adstock decay rates. We implement the approach using NumPyro (Phan et al. 2019) with JAX acceleration, which recent benchmarks show achieves 2-20x faster sampling than TensorFlow-based alternatives while maintaining superior channel contribution recovery (PyMC Labs 2025).

2. Model Specification

Let x_t denote marketing spend at time t and y_t the observed outcome. The model proceeds as follows.

Adstock transformation. We apply geometric decay to capture carryover effects:

$$s_t = x_t + \alpha \cdot s_{t-1}, \quad s_0 = 0$$

where $\alpha \in [0, 1]$ is the decay parameter with prior $\alpha \sim \text{Beta}(2, 2)$.

Hill saturation. For each latent segment $k \in \{1, \dots, K\}$, the response function is:

$$f_k(s) = A_k \cdot \frac{s^{n_k}}{\lambda_k^{n_k} + s^{n_k}}$$

where A_k is the maximum effect, λ_k is the half-saturation point, and $n_k \sim \text{LogNormal}(\log 1.5, 0.4)$ controls curve steepness.

Mixture likelihood. The observation model is a Gaussian mixture:

$$y_t \sim \sum_{k=1}^K \pi_k \cdot \mathcal{N}(\mu_0 + \beta t + f_k(s_t), \sigma^2)$$

where $\pi \sim \text{Dirichlet}(\mathbf{1}_K)$ and μ_0, β capture baseline trend.

Identifiability. Mixture models suffer from label switching, where posterior samples may exchange component labels across MCMC iterations. We impose ordering via cumulative sum reparameterization: $\lambda_k = \sum_{j=1}^k \delta_j$ with $\delta_j \sim \text{LogNormal}(\log(s_{\max}/(K+1)), 0.7)$, ensuring $\lambda_1 < \lambda_2 < \dots < \lambda_K$.

Automatic prior scaling. To accommodate diverse data scales, priors are computed automatically from training data: $A_k \sim \text{LogNormal}(\log(0.3 \cdot \text{range}(y)), 0.8)$, $\mu_0 \sim \mathcal{N}(\bar{y}, 2\sigma_y)$, and $\sigma \sim \text{HalfNormal}(\sigma_y)$.

3. Experiments

Data generation. We evaluate on simulated data with known ground truth ($T = 200$ observations). The data-generating process uses $K = 3$ true components with: mixing proportions $\pi = (0.40, 0.30, 0.30)$; half-saturation points $\lambda = (0.5s_{\text{med}}, s_{\text{med}}, 1.2s_{\text{med}})$; maximum effects $A = (15, 30, 60)$; Hill coefficients $n = (2.0, 1.5, 1.0)$; adstock decay $\alpha = 0.5$; baseline intercept $\mu_0 = 50$ and slope $\beta = 2$; and observation noise $\sigma = 3$.

Model comparison. We compare three specifications using a 150/50 train/test split: (1) **Single Hill** with one global saturation curve, (2) **Mixture K=3** matching the true number of components, and (3) **Sparse K=5** with Dirichlet concentration 0.5 for automatic pruning. Inference uses NUTS with 1,000 warmup and 3,000 sampling iterations across 4 chains.

Results. All models achieved $\hat{R} < 1.01$ and $\text{ESS} > 400$, indicating proper convergence. Table 1 summarizes predictive performance:

Model	Train RMSE	ELPD-LOO	Test RMSE	α (true: 0.5)
Single Hill	4.12	-412.5	4.28	0.48 ± 0.06
Mixture K=3	3.24	-389.7	3.41	0.51 ± 0.05
Sparse K=5	3.31	-394.2	3.52	0.49 ± 0.05

The Mixture K=3 model achieved a 22.8 point improvement in ELPD-LOO over Single Hill (SE = 8.4), indicating substantially better out-of-sample predictive accuracy. The adstock parameter

α was accurately recovered across all models, with 95% credible intervals containing the true value 0.5.

4. Discussion and Conclusion

The proposed Hill mixture model addresses a fundamental limitation of standard MMM by allowing heterogeneous consumer response. Key advantages include: (1) interpretable segment-specific parameters, (2) full uncertainty quantification through Bayesian inference, and (3) automatic complexity control via sparse Dirichlet priors.

Limitations. The approach assumes segment membership is constant over time, which may not hold during product launches or competitive shifts. The mixture structure also increases computational cost compared to single-curve models.

We presented a Bayesian mixture of Hill saturation functions for capturing heterogeneous consumer response in Marketing Mix Modeling. The approach achieves superior predictive performance compared to single-curve baselines while maintaining interpretability through segment-specific response curves. Implementation in NumPyro with automatic prior scaling provides a practical tool for practitioners.

References

- Allenby, Greg M., and Peter E. Rossi. 1998. “Marketing Models of Consumer Heterogeneity.” *Journal of Econometrics* 89 (1-2): 57–78.
- Chan, David, and Mike Perry. 2017. “Challenges and Opportunities in Media Mix Modeling.” *Google Research White Paper*.
- Dew, Ryan, Nicolas Padilla, and Anya Shchetkina. 2024. “Your MMM Is Broken: Identification of Nonlinear and Time-Varying Effects in Marketing Mix Models.” *arXiv Preprint arXiv:2408.07678*.
- Jin, Yuehan, Yuxue Wang, Yunting Sun, David Chan, and Jim Koehler. 2017. “Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects.” *Google Research White Paper*.
- Phan, Du, Neeraj Pradhan, and Martin Jankowiak. 2019. “Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro.” *arXiv Preprint arXiv:1912.11554*.
- PyMC Labs. 2025. *PyMC-Marketing Vs. Meridian: A Quantitative Comparison of Open Source MMM Libraries*. PyMC Labs.
- Wedel, Michel, and Wagner A. Kamakura. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. 2nd ed. Springer Science & Business Media.