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

Gregory Szep

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. Comprehensive experiments on simulated data across varying complexity levels demonstrate that mixture models achieve superior predictive accuracy compared to single-curve baselines when the true data-generating process is heterogeneous, with the sparse mixture approach offering better convergence properties.

**Keywords:** Marketing Mix Modeling, Bayesian inference, Hill function, mixture models

## 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:

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

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

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

$$f_k(s) = A_k \cdot \frac{s^{n_k}}{\lambda_k^{n_k} + s^{n_k}} \tag{2}$$

where $A_k$ is the maximum effect, $\lambda_k$ 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:

$$y_t \sim \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(\mu_0 + \beta t + f_k(s_t), \, \sigma^2\right) \tag{3}$$

where $\pi \sim \text{Dirichlet}(\mathbf{1}_K)$ and $\mu_0, \beta$ 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 $\lambda_1 < \lambda_2 < \cdots < \lambda_K$.

**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 model performance across four data-generating processes (DGPs) with varying complexity: (1) **Single** ($K = 1$): standard single Hill response, (2) **Mixture** $K = 2$: two-component mixture, (3) **Mixture** $K = 3$: three-component mixture, and (4) **Mixture** $K = 5$: five-component mixture. Each DGP uses $T = 200$ observations with a 150/50 train/test split. We run 5 random seeds per condition to assess variability.

**Model comparison.** We compare three specifications: (1) **Single Hill** with one global saturation curve, (2) **Mixture** $K = 3$ with three components, and (3) **Sparse** $K = 5$ with five components and Dirichlet concentration 0.5 for automatic pruning. Inference uses NUTS with 1,000 warmup and 3,000 sampling iterations across 4 chains.

**Results.** Table 1 summarizes predictive performance using expected log pointwise predictive density (ELPD-LOO) via leave-one-out cross-validation.

Key findings: (1) When the true DGP is simple ($K = 1$), mixture models incur only marginal ELPD penalty (1–2 points), avoiding substantial overfitting. (2) For heterogeneous data ($K \geq 2$), mixture models achieve significant ELPD improvements, with the largest gain of 32.7 points observed at $K = 3$. (3) Test RMSE is similar across models within each DGP, indicating that ELPD improvements reflect better uncertainty quantification rather than point prediction accuracy.

Table 1: Predictive performance across data-generating processes and model specifications.

| True $K$ | Model | ELPD-LOO | Test RMSE | $\Delta$ LOO vs Single |
|---|---|---|---|---|
| $K = 1$ | Single Hill | $-382.4 \pm 11.9$ | $5.1 \pm 0.5$ | — |
| $K = 1$ | Mixture $K = 3$ | $-384.0 \pm 11.7$ | $5.1 \pm 0.5$ | $-1.6$ |
| $K = 1$ | Sparse $K = 5$ | $-384.4 \pm 11.6$ | $5.2 \pm 0.5$ | $-2.0$ |
| $K = 2$ | Single Hill | $-409.7 \pm 8.6$ | $5.4 \pm 0.5$ | — |
| $K = 2$ | Mixture $K = 3$ | $-406.3 \pm 8.6$ | $5.4 \pm 0.6$ | $+3.5$ |
| $K = 2$ | Sparse $K = 5$ | $-406.9 \pm 8.4$ | $5.4 \pm 0.5$ | $+2.8$ |
| $K = 3$ | Single Hill | $-520.7 \pm 7.4$ | $8.3 \pm 1.0$ | — |
| $K = 3$ | Mixture $K = 3$ | $-488.0 \pm 9.5$ | $8.4 \pm 0.9$ | $+\mathbf{32.7}$ |
| $K = 3$ | Sparse $K = 5$ | $-488.1 \pm 9.3$ | $8.4 \pm 0.9$ | $+\mathbf{32.7}$ |
| $K = 5$ | Single Hill | $-509.5 \pm 4.7$ | $7.8 \pm 0.9$ | — |
| $K = 5$ | Mixture $K = 3$ | $-498.4 \pm 5.4$ | $7.8 \pm 0.9$ | $+11.1$ |
| $K = 5$ | Sparse $K = 5$ | $-498.0 \pm 5.5$ | $7.8 \pm 0.9$ | $+11.5$ |

**Convergence diagnostics.** The Sparse $K = 5$ model demonstrates superior convergence properties. For the $K = 3$ DGP, the Mixture $K = 3$ model exhibited convergence difficulties (0/5 runs achieved $\hat{R} < 1.05$, with $\hat{R}$ ranging from 1.22 to 1.78), while Sparse $K = 5$ achieved convergence in 3/5 runs. This suggests that the additional flexibility and sparsity-inducing prior of the Sparse $K = 5$ specification helps avoid local modes.

**Effective component recovery.** The sparsity mechanism successfully recovers approximate complexity: for $K = 1$, effective components $\approx 1.9$; for $K = 2$, $\approx 2.6$–$2.8$; for $K = 3$, $\approx 2.9$–$3.7$. However, when true $K = 5$, both models underestimate complexity (effective $K \approx 3.5$–$3.7$), suggesting limitations in recovering high-dimensional mixtures.

## 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 increases computational cost compared to single-curve models. Additionally, our experiments revealed that the Mixture $K = 3$ specification can suffer from convergence difficulties when the true complexity matches or exceeds the model capacity. The Sparse $K = 5$ model, while more robust, may underestimate the true number of components when heterogeneity is high.

**Recommendations.** Based on our experimental findings, we recommend the **Sparse $K = 5$** specification as the default choice for practitioners. It adapts well to both simple ($K = 1$) and complex ($K \geq 2$) scenarios while offering substantially better convergence properties than the fixed

Mixture $K = 3$ model. The sparsity-inducing Dirichlet prior prevents overfitting when the true structure is simple.

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 when heterogeneity is present, while maintaining competitive performance on homogeneous data. Implementation in NumPyro with automatic prior scaling provides a practical tool for practitioners.

# References

Greg M. Allenby and Peter E. Rossi. Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89(1-2):57–78, 1998.

David Chan and Mike Perry. Challenges and opportunities in media mix modeling. *Google Research White Paper*, 2017.

Ryan Dew, Nicolas Padilla, and Anya Shchetkina. Your MMM is broken: Identification of nonlinear and time-varying effects in marketing mix models. *arXiv preprint arXiv:2408.07678*, 2024.

Yuehan Jin, Yuxue Wang, Yunting Sun, David Chan, and Jim Koehler. Bayesian methods for media mix modeling with carryover and shape effects. *Google Research White Paper*, 2017.

Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable effects for flexible and accelerated probabilistic programming in NumPyro. *arXiv preprint arXiv:1912.11554*, 2019.

PyMC Labs. PyMC-Marketing vs. Meridian: A quantitative comparison of open source MMM libraries. Technical report, PyMC Labs, September 2025.

Michel Wedel and Wagner A. Kamakura. *Market Segmentation: Conceptual and Methodological Foundations*. Springer Science & Business Media, 2nd edition, 2000.